Current Readings in Educational Psychology

SECOND EDITION

OTHER BOOKS BY THE AUTHOR

The Psychology of College Success: A Dynamic Approach

Psychology: An Introduction to a Behavioral Science, 3rd edition
with Donn Byrne

Current Research in Psychology: A Book of Readings
edited with Donn Byrne & Fredrica Lindgren

Study Guide to Accompany: Psychology: An Introduction to a Behavioral Science
with Donn Byrne

Educational Psychology in the Classroom, 4th edition

Study Guide to Accompany Educational Psychology in the Classroom

An Introduction to Social Psychology

Current Readings in Educational Psychology

SECOND EDITION

EDITED BY

Henry Clay Lindgren

Professor of Psychology
San Francisco State College

and

Fredrica Lindgren

John Wiley & Sons, Inc.

New York London Sydney Toronto

Copyright © 1968, 1971 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means, nor transmitted, nor translated into a machine, language without the written permission of the publisher.

Library of Congress Catalog Card Number: 71-168638
ISBN 0-471-53651-2 (paper)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

One of the continuing and perplexing problems facing an instructor in educational psychology courses is that of finding ways to introduce his students to interesting and relevant reading material that will take them beyond the confines of their textbooks. Even the best textbooks, of necessity, must be synoptic and even somewhat superficial when it comes to doing justice to more than a limited selection of the many excellent research studies and theoretical papers that are potentially valuable sources of information, ideas, and integrating concepts.

The solution to this problem appears to be found in books of readings. The fact that instructors are turning to them more and more shows that these books do fill a real instructional need. Not only do students get a chance to become familiar with the work of leaders in the field but many an article that, otherwise, might be read (at most) by a few hundred scholars is made available to thousands of students. Both the student and the author of the journal article gain by the increased exposure that the book of readings provides.

All books of readings are unavoidably samplers, and they obviously reflect the biases of their editors. A review of the selections in this sampler probably would reveal my continuing interest in the emotional and social aspects of classroom learning, although I have gone to some lengths to see that other viewpoints and positions are represented as well.

The general motive behind our selection of articles is to provide students in educational psychology courses with a sampling of the kind of reading material that interests educational psychologists. The papers in this book include theory, opinion, research techniques and findings, comments, and methods. Although most of the selections were written by educational psychologists, the contributors also include psychologists from clinical, social, developmental, industrial, and counseling fields, as well as sociologists, psychiatrists, and even a biologist.

The sequence of the 51 readings follows the general order of topics presented in the senior editor's textbook, *Educational Psychology in the Classroom*, 4th edition (Wiley, 1972). Part 1 consists of a historical review of the development of educational psychology as an applied behavioral science, as well as a discussion of the statistical concepts, terms, and processes that students will encounter in some of the papers included in the book. Part 2 presents concepts and research data that may aid in helping teachers develop an

understanding of the learner in his psychosocial environment. Part 3 deals specifically with the learner's mental health —always a factor in classroom learning. Part 4 is concerned with the learning process: factors affecting learning, self-theories of learning, learning through discovery, and applications of learning theory to instruction. The learning situation is the man concern of Part 5: behavior modification, and classroom management are the major topics. Four discussions of the evaluation of learning are presented in Part 6, while Part 7 presents papers on intelligence and creativity, including two on the controversy regarding the genetic factors in intelligence. The papers in Part 8 are concerned with the individualization of education: special treatment for students with special problems. This theme is continued in Part 9, which consists of a group of papers dealing with the problems of the socially disadvantaged child. The selections in Part 10 deal with the psychology of the teacher.

As authors-editors of this collection of papers, we would like to express our appreciation and thanks to the contributors and publishers, who permitted these articles to be published here.

Henry Clay Lindgren
Fredrica Lindgren

San Francisco

Contents

PART 1 Introduction

1.1 *A Brief History of Educational Psychology* 5
ROBERT I. WATSON

1.2 *Interpreting Research in Educational Psychology* 34
DONN BYRNE AND HENRY CLAY LINDGREN

PART 2 Growth and Development of Learners

2.1 *The Beginning of the Self: The Problem of the Nurturing Environment* 55
IRA J. GORDON

2.2 *Parental Expectations for Independent Behaviors and Achievement of Elementary School Boys* 62
ANITA WHITING

2.3 *Children's Self-Concept and Kindergarten Achievement* 70
RICHARD J. OZEHOSKY AND EDWARD T. CLARK

2.4 *A Study of Seventh Grade Children's Reading of Comic Books as Related to Certain Other Variables* 77
W. PAUL BLAKELY

2.5 *Where Are the Siblings? A Re-evaluation of the Relationship Between Birth Order and College Attendance* 88
WILLIAM T. SMELSER AND LOUIS H. STEWART

PART 3 Mental Health of Learners

3.1 *Interpersonal Relations and Mental Health in the Classroom* 101
RICHARD A. SCHMUCK, MARGARET B. LUSZKI, AND DAVID C. EPPERSON

3.2 *Social Schema of Normal and Disturbed School Children* 114
RHODA LEE FISHER

3.3 *The Disturbed Child in the Classroom* 121
KATHARINE F. TIFT

3.4 *Effects on Cheating of Achievement Anxiety and Knowledge of Peer Performance* 126
JEV SHELTON AND JOHN P. HILL

3.5 *Alienation of Present-Day Adolescents* 131
LOUIS J. WISE, M.D.

3.6 *Reducing Educational Pressures* 143
RICHARD DEMPSEY

PART 4 The Learning Process: Theory, Research, and Practice

4.1 *Psychoneurobiochemeducation* 149
DAVID KRECH

4.2 *A Self-Concept Theory of Learning: A Learning Theory for Teachers* 160
WALCOTT H. BEATTY AND RODNEY CLARK

4.3 *The Act of Discovery* 182
JEROME S. BRUNER

4.4 *The Motivating Effect of Learning by Directed Discovery* 195
BERT Y. KERSH

4.5 *A Comparative Study of Three Preschool Curricula* 203
DAVID P. WEIKART

4.6 *Modern Learning Theory and the Elementary School Curriculum* 216
PATRICK SUPPES

PART 5 The Learning Situation and Its Management

5.1 *Human Motivation and Environment* 235
LLOYD E. HOMME

5.2 *Negative-Attention Seeking or Tactical Provocation?* 246
RONALD GALLIMORE

5.3 *Social Reinforcement in Reducing Inappropriate Behavior* 253
ALAN S. BRISKIN AND WILLIAM I. GARDNER

5.4 *The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct* 261
JACOB S. KOUNIN AND PAUL V. GUMP

5.5 *Changes in Pupil Attitudes During the School Year* 270
NED A. FLANDERS, BETTY M. MORRISON, AND E. LELAND BRODE

5.6 *What Are We Rewarding?* 278
JAMES A. CONWAY

PART 6 The Evaluation of Learning

6.1 *The Role of Evaluative Research* 287
EDWARD A. SUCHMAN

6.2 *The Functions and Uses of Educational Measurement* 296
WINTON H. MANNING

6.3 *Testing Programs—Misconception, Misuse, Overuse* 306
FRANK B. WOMER

6.4 *Improving the Competence of Teachers in Educational Measurement* 314
ROBERT L. EBEL

PART 7 Individual Differences: Intelligence and Creativity

7.1 *Heredity and Environment: A Controversy over IQ and Scholastic Achievement* 323
ARTHUR R. JENSEN

7.2 *Educational Relevance and Jensen's Conclusions* 328
NICHOLAS ANASTASIOW

7.3 *Diversity* 336
JOHN R. PLATT

7.4 *Brainstorming and Orneriness as Facilitators of Creativity* 344
HENRY CLAY LINDGREN AND FREDRICA LINDGREN

7.5 *Training Children to be Creative May Have Little Effect on Original Classroom Performance* 352
FRANK E. WILLIAMS

PART 8 Providing for Individual Differences

8.1 *Effects of Accelerating Bright, Older Elementary Pupils—a Second Follow-up* 363
HERBERT J. KLAUSMEIER, WILLIAM L. GOODWIN, AND TECKLA RONDA

8.2 *The Ombudsman: Relevant Role Model for the Counselor* 373
MICHAEL A. CIAVARELLA AND LAWRENCE W. DOOLITTLE

8.3 *A Work Experience Program for the Educable Mentally Retarded at the Elementary School Level* 380
MARY TURNER

8.4 *Educational Disability in Emotionally Disturbed Children* 384
F. BETH STONE AND VINTON N. ROWLEY

8.5 *School Grades and Group Therapy* 390
SAMUEL TENENBAUM

8.6 *New Programs for Dropouts* 397
HERMAN SLOTKIN

PART 9 The Socially Disadvantaged Learner

9.1 *Malnutrition and Learning* 409
RITA BAKAN

9.2 *Mother–Child Interaction: Social Class Differences in the First Year of Life* 418
STEVEN R. TULKIN AND JEROME KAGAN

9.3 *Measures of Intelligence on Southwest Indian Students* 422
BERT P. CUNDICK

9.4 *The Early Training Project: A Seventh Year Report* 428
SUSAN W. GRAY AND RUPERT A. KLAUS

9.5 *A Comparison of the Problems of Certain Anglo- and Latin-American Junior High School Students* 441
ROBERT PAUL WITHERSPOON

9.6 *Negro Academic Motivation and Scholastic Achievement* 447
ROBERT LEE GREEN AND WILLIAM W. FARQUHAR

9.7 *Changes in Attitudes Toward Negroes of White Elementary School Students After Use of Multiethnic Readers* 452
JOHN H. LITCHER AND DAVID W. JOHNSON

PART 10 The Psychology of Being a Teacher

10.1 *Teachers' Communication of Differential Expectations for Children's Classroom Performances: Some Behavioral Data* 463
JERE E. BROPHY AND THOMAS L. GOOD

10.2 *The American Teacher: A Tentative Psychological Description* 480
LOUIS S. LEVINE

10.3 *Authoritarianism, Independence, and Child-Centered Practices in Education: A Study of Values* 491
HENRY CLAY LINDGREN

10.4 *Anxieties and Discontents in Teaching* 495
ROBERT H. SNOW

INDEX **503**

Current Readings in Educational Psychology

SECOND EDITION

Ken Hyman

PART 1

Introduction

1.1 *A Brief History of Educational Psychology*

ROBERT I. WATSON

Reprinted with slight abridgment from the *Psychological Record, 1961,* **11**, 209–242, with permission of the author and the *Psychological Record.* For complete source citations, see the original article. Robert I. Watson is professor of psychology at the University of New Hampshire and has written extensively on clinical and developmental psychology, as well as on the history of psychology as a science.

In this opening selection, Robert I. Watson provides us with an overview of educational psychology from its very beginnings with the Greek philosophers up to World War II. The student who expects to continue his work in the field of educational psychology will appreciate Watson's detailed description of the early trends and issues that led to the differentiation of educational psychology from general psychology. Students who expect to become classroom teachers, however, will probably be more interested in the broader and more general themes developed by Watson and in learning how a large number of eminent scholars and scientists, whose names they have already encountered in other contexts, have contributed to the development of educational psychology as a major applied field within the context of psychology.

Before Educational Psychology

Before the emergence of psychology as a discipline in its own right, and of educational psychology as a branch of that discipline, speculation and observation concerning the relation of human nature to the educational process were not uncommon. This period began in classical Greece. If one examines the fragments of the thinking of the pre-Socratic philosophers that have been preserved there is to be found a variety of comments relevant to this relation. Democritus, who was in his prime early in the fifth century B.C., will serve as an example. He wrote not only about the advantages of education but also considered the influence of the home upon the child. He held that self-control of the father served to teach the children. Stress was laid upon such matters as training in the management of property by sharing it with the children. He also commented that some qualities present in the parents result in their opposite appearing in the child, giving as an instance the parental practice of excessive thrift leading to extravagance in the children.

Of Plato and of Aristotle, who flourished in the fourth century B.C., we have much fuller accounts of their views of education and its relation to psychological factors. The scope of their thinking might seem to some persons today to be surprisingly complex and broad. Scattered throughout their works Aristotle and Plato discuss the following: the ends of education; the ideal of the educated man, the disadvantages of being educated; the kinds of education that are appropriate to different kinds of people; the training of the body, and the cultivation of bodily skills; the formation of a good character; the possibilities and limits of moral education; the influence of the family in this training; the role of the state in moral education; the effect upon character of music, poetry and the other arts; the profession of teaching, and the relation of teacher and student; the means and methods of teaching; the nature of learning; the order of learning; the emotional aspects of learning; learning apart from teachers, and the acquisition of techniques.

Aristotle's psychological views relevant to educational matters are presented more systematically, less fancifully, and in greater detail than are those of Plato. Two aspects of Aristotle's thinking especially worthy of note are that his writings served to encourage a faculty psychology and that he emphasized the intellectual processes. Aristotle laid the foundation for the main psychological doctrines taught and accepted by the learned world for the next 2000 years. In the earlier darker years such knowledge virtually disappeared, only to be revived again.

Throughout the centuries, the formulations of Aristotle have been modified in particulars. The most noteworthy of these modifications took place when the medieval scholastic thinkers attempted to reconcile Christian doctrine with Aristotelian teaching. These efforts reached their climax in the work of Thomas Aquinas in the thirteenth century.

Descartes in the seventeenth century introduced powerful, reasoned support for the doctrine that ideas which are innate, rather than those arising from ex-

perience, are the basis of true knowledge.

It was John Locke, a seventeenth century empiricist, who initiated the first wide and continued protest against the form of faculty psychology current in his time. He did not abandon faculties altogether, but argued that the faculties were not real things in the soul that performed the actions indicated by their various names. Locke's *Essay concerning Human Understanding* (first published in 1690 and including in the fourth edition, published in 1700, a chapter entitled "Of the Association of Ideas") is an attempt to combat the Cartesian theory of innate ideas. Locke insisted that, at birth, the human mind was not performed and ready to function, but potentially sensitive to impressions from the external world through the senses. Through the empirical process of experience he receives sensory experiences from which simple experiences are built into complex ideas through the internal activity of reflection. Ideas, then, were not already present but arose from sensory impressions and from the process of reflection. Thus arise all knowledge, all values, and all morality. This learning through experience came to be known as empiricism. Other psychologists that came after him took the position that the source of ideas was sensory experiences alone and rejected his faculty of reflection. From his work and from the work of those that followed after, a thoroughgoing nonfaculty point of view began to emerge—association psychology. Ironically enough, this associationist point of view also goes back to Aristotle who discussed the lawfulness of association and spoke of the importance of similarity, contrast, and contiguity. The struggle over faculty psychology and over association psychology, eventually sharply separated, continued for about two hundred years after Locke's publications at the turn of the eighteenth century.

In the course of the development of the doctrine of faculties its ramifications, subtleties and nuances were as detailed, contradictory and as hotly contested as any of the ideas of Western man. The doctrine may be simplified to the statement that most commonly the mind was considered as three interdependent (or, conversely, dependent), sets of powers or capacities, the first and foremost of which was understanding, reasoning or intellect, another was feelings, desires, emotions and appetites, and the third was the will or volition.

At the beginning of the nineteenth century, although continuing the faculty psychology, Pestalozzi changed the sphere of teacher training by placing an emphasis upon education as a process of drawing out of the individual. This, in turn, required a method to be sought to minimize that of the memorizing and the hearing of recitations which was then prevalent. Proceeding from the concrete to the abstract, attending to the harmoniousness of development, and searching for the laws of development characterized his point of view.

Faculty psychology held an unquestioned firm grip upon education in the United States from the 1830's to almost the end of the last century. Faculty psychology set the limits and form of education. It provided the basic justification for the method of mental discipline, giving first place to a liberal education, in the original sense of the term, and

served as an obstacle to the appearance of new and potentially useful subjects in the curriculum. The faculties of the students were to be exercised by use of those subjects of instruction, which, in the words of Day, President of Yale in the early decades of the century, are "best calculated to teach the art of fixing the attention, directing the train of thought, analyzing a subject proposed for investigation, following, with accurate discrimination, the course of argument; balancing nicely the evidence presented to the judgment; awakening, elevating, and controlling the imagination; arranging, with skill, the treasures which memory gathers; rousing and guiding the powers of genius. . . ." The concentration on the intellectual faculties to the virtual exclusion of the faculties of feeling and will is obvious.

The next great advance in educational psychology came about after the middle part of the century when the neglected work of Johann Friedrich Herbart became influential. Herbart was a German professor, who, according to Eby and Arrowood, was the first to formulate an approach to education based directly and avowedly upon psychology. However, as his predecessors before him, he was primarily philosophical in aim, content, and method. He specifically denied the possibility of experimenting upon the mind. Nevertheless, he rejected the faculty psychology of his predecessors and stressed the importance of interest and apperception and apperceptive mass. Herbart not only opposed the faculty psychologists of his time, but also considered the human personality as a dynamic and individually structured system of forces. Ideas are active and struggle with one another for a place in consciousness. His work in education was carried on adjunct to his major task of philosophy, and at the time of his death in 1841 and for some years thereafter went practically unnoticed. In 1868 a professor at Leipzig, Ziller, published a book calling attention to his work. This book was widely read. Herbart's influence spread rapidly thereafter with his ideas elaborated and advanced as the primary basis for school instruction. In the 1890's the influence of Herbart upon educational psychology in the United States was all-embracing in nature.

In England, meanwhile, two important books were published. The volume of Alexander Bain, *Education as a Science,* was published in 1878 and emphasized an associationist view. James Sully's *Outlines of Psychology with Special Reference to the Theory of Education,* was published in 1884 and emphasized a faculty point of view. Because of the all-embracing nature of these conflicting two views at that time, it is quite fitting that mention of these two contrasting books should close the account of the period preceding the advent of educational psychology as an aspect of psychology.

The Beginnings of Educational Psychology: 1880–1900

There is considerable justification for placing the beginnings of educational psychology in the early years of the 1880's. During this time Galton published the first experimental investigation of associationism, and began giving to the public a series of physical mea-

surements and tests of reaction time and sensory acuity. Hall, meanwhile, published the first of his papers using the questionnaire to investigate the minds of children. In 1885 Ebbinghaus published his study of memory, *Ueber das Gedächtnis.* In the span of six years events important in the founding of objective measurement, child psychology, and learning all took place. In the present perspective they are also seen to be crucial aspects of educational psychology.

Francis Galton

Sir Francis Galton, the oldest of the founders of educational psychology, was influenced by the evolutionary associationist tradition, particularly the views espoused by Lewes and Spencer. Associationism under the influence of Darwin had taken on a more developmental cast. As was just mentioned, it was Galton who published in 1879 the first experimental investigations of associationism. The three studies he reported in this paper were concerned with the free play of association, the reaction time necessary to produce associations, and perseveration, or the tendency to repeat the same associations. His work on association was repeated in Wundt's laboratory the following year. Seven years later, also at the Leipzig laboratory, James McKeen Cattell performed his well known series of studies on reaction time in relation to various mental processes. However, these contributions from Wundt's pupils had only an indirect connection with further developments in associationism.

To return to Galton, his work was guided by two basic Darwinian assumptions—that there is a resemblance in innate constitution of offspring to their parents, and that the innate constitution of the offspring would, because of spontaneous variation, bring about to some degree differences both from their parents and from one another. In 1869 he published his study, *Hereditary Genius,* which showed that eminent men in England tend to be related to one another. He conducted many other studies on mental inheritance among persons of ability and studied the resemblance of twins. Galton became firmly convinced of the importance of hereditary differences. With this emphasis upon hereditary differences, a wide-spread realization of the importance of individual differences, in the psychological sense, came into being. This interest in the nature-nurture problem took the form not only of individual differences conceived as being innate, but also seen as being due to environmental differences. Galton's enthusiasm for problems in this area led to his making an important advance in the history of statistics—the idea of correlation. This was followed up by Karl Pearson who developed the measure of relationships that bears his name. The concept was that of Galton; the technique was that of Pearson. Further contributions from Galton took the form of the development and use of psychological tests, rating scales and questionnaires. His most important theoretical contribution was the distinction in the "structure of mind" between a general broad ability, or intelligence, and special abilities entering only into narrower ranges of activity. His conceptual distinction, his statistical tool

and his testing devices bore rich fruit in studies of factor analysis.

G. Stanley Hall

Although the influence of G. Stanley Hall has been most directly felt in child psychology, he deserves mention in a discussion of the history of educational psychology. From the 1880's onward as a well known psychologist, as an academic leader in his capacity of President of Clark University, and as a promoter of worthy causes in psychology he exerted an influence which overrode considerably the theoretical weaknesses inherent in his own research. His conceptual scheme was based upon theories of biological evolution and recapitulation, in which the mind was seen as growing through a series of stages, more or less corresponding to those through which it was alleged early man had gone. Although enjoying considerable vogue at the time, today there remains very little trace of this point of view.

He became interested in the possibilities of the questionnaire which he had found being used earlier in Berlin during the 1870's. With his characteristic flair for creating enthusiasm in others he inspired much research designed to discover the contents of children's minds. In the next decade his first paper appeared that used the questionnaire for the purpose. This was followed by a whole series of studies, some carried out by him and his students but more carried out by others. The use of the technique became a fad, being applied by parents, teachers and others and a rash of uncritical, poorly designed, superficial, statistically unsophisticated papers appeared. This use of the questionnaire was the major impetus for the development of the so-called child study movement, with societies formed here and abroad. Hall, himself, lost interest in the movement before it had run its course, but through his inspiration and through his providing, beginning in 1891, a vehicle for publication, in the *Pedagogical Seminary* (now the *Journal of Genetic Psychology*) he was the major impetus for the movement. The period of most intense activity of this movement was from about 1890 to 1915. Eventually, the child study movement, in its original sense, fell of its own weight, yet the movement made its positive contributions in the increased recognition of the importance of the empirical study of the child, an increased critical evaluation of research, and a recognition of the importance of the study of childhood development itself.

William James

William James, to use Boring's expression (1950), was a "self starter." Influenced to some degree by many individuals he had no mentor but himself. His book, *Principles of Psychology,* published in 1890, had a profound effect upon the development of psychology as a natural science. Drawing upon evolutionary biology, physiology, associationism, and all of the other scattered intellectual current of his day, in this book he carved out the beginnings of the so-called functional approach to psychology. In it, he saw the mind in terms of use in survival and in competition. Critical of the crudities of materialistic physiological psychologists and of the

brass instrument psychology, and with a vivid, even epigrammatic, style he did much to win from teachers and others an appreciation of what, to a considerable extent, had been heretofore generally regarded as a curiously lifeless subject—the mind. The book, itself, is free of specific reference to educational applications. Rather, it served as a source for others to draw implications for education. It performed what Morris (1950) so aptly called its particular function, that of a watershed, initiating the flow of the future. One of the lessons to be drawn from Allport's discussion (1943) of the paradoxical (and often contradictory) positions that James took on psychological issues is that his work is a gigantic projective reservoir of ideas. One goes to James for support on a matter of concern to psychology and finds it put forth enthusiastically and convincingly. If he does not look for contrary pronouncements, disregards them, disagrees with them, or harmonizes them, he can come away refreshed and supported. James also contributed to educational psychology through his ability as a public speaker and popularizer. Despite its mundane title, his *Talks to Teachers,* published in 1899, is a brilliant book, and is crammed with still pertinent practical suggestions on educational matters given in his usual vivid style.

James McKeen Cattell

The early work of James McKeen Cattell on individual differences in reaction time in Wundt's laboratory has already been mentioned. This research, recognized as not being in the tradition of the Leipzig Laboratory, was prophetic of Cattell's later attitude. He broke away from the Wundtian outlook in that he laid stress on individual differences, and vigorously defended the practical implications of psychological findings. His student days at Leipzig are so much a part of the folklore of psychology that it is sometimes overlooked that he was thereafter a lecturer at Cambridge University where he was much influenced by the man he said was "the greatest man whom I have known," Francis Galton. Cattell, on his return from Germany, served as Professor of Psychology at the University of Pennsylvania and then moved to Columbia University where he was head of the Laboratory and of the Department for 26 of the years in which psychology was shaping itself.

Much of his influence on educational psychology was of a more general character than was that of some of the other pioneers. He worked assiduously toward promoting all applications of psychology, including educational psychology. He organized journals (along with J. Mark Baldwin) (*The Psychological Review, Psychological Monographs,* and *Psychological Index*), served as an editor *(Science, American Men of Science, Scientific Monthly,* and *School and Society*), and was a founder of the Psychological Corporation.

It was in the area of mental tests that he made his most direct contribution to educational psychology. While still at the University of Pennsylvania, he had administered mental tests to the students and, indeed, in describing the tests, coined the term, "mental tests." He continued his testing program when he moved to Columbia. Data were collected for several years and a monograph on

the results was prepared by Clark Wissler (1901). The correlations between the tests and academic class standing were hardly better than zero. Moreover, the tests did not correlate appreciably better among themselves. This was in sharp contrast to the substantial correlations among the academic subjects. The results, then, seemed to lead nowhere so far as the usefulness of the tests was concerned. An earlier study by Sharp (1899) had appeared from Titchener's laboratory at Cornell University with similar negative results. These two studies did much to make psychologists lose interest in the topic for the next few years. The tests that had been used were laboratory sensory and motor tasks—memory for ideas, association, memory span, reaction time, perceptual discrimination, and the like. They were not organized into scales but each test was treated separately. At the time it was not appreciated that these devices were not suitable as educational measures. It was not until Americans became familiar with the work of Alfred Binet that interest in mental testing revived.

Alfred Binet

The next major contributor to the founding of educational psychology is Alfred Binet, who developed the first widely used individual intelligence test. Binet introduced objectivity into an area which, before his work, had lent itself poorly to direct vigorous investigation. In 1890 he had become associated with the Sorbonne and its laboratory of physiological psychology, an association which continued until his death in 1911. At first he had worked on problems of abnormal psychology, but from about 1887 onwards his research was conducted primarily in the schools of Paris and its suburbs. In collaboration with V. Henri, he had contended in the very first issue of *L'Année Psychologique,* in 1895, that individual psychology was much more significantly studied by investigation of the complex mental processes than by tests of sensory discrimination and reaction tests. Pursuing this aim in the decade that followed this paper, Binet carried on many studies of the complex processes, such as tests of memory of words and of sentences, descriptions of objects and of pictures, and attention. The specific task that set him upon the development of the Binet Scale was the necessity of having an instrument whereby Parisian school children suspected of mental retardation could be identified so that they might be given special training. With the assistance of Théophile Simon he developed the first Binet Scale.

When Binet and Simon were working on their first Scale there was sufficient work already accomplished to lead them to refer to the widespread use of tests. The fundamental concept they followed in its construction was that of what Binet called a metrical scale—tests with items arranged in order of increasing difficulty, with each test corresponding to a different mental level. Deliberately they selected a considerable variety of tests in order to tap various forms of mental capacity. Among the 30 tests in the 1905 scale there were measures of visual coordination, execution of simple orders, knowledge of objects, repetition of sen-

tences, giving differences between pairs of familiar objects, repetition of digits, and making distinctions between abstract forms. They arranged the tests, on the basis of preliminary findings, in an ascending order of difficulty and administered the tentative scale to normal and subnormal children. Norms for children of ages three, five, seven, nine, and eleven were derived from this. The concept of mental age was clearly grasped but not worked out in practice until their 1908 and 1911 scales which further refined standardization of the instrument. (The introduction of the Binet-Simon Scale in the United States is considered later.)

John Dewey

The sheer diversity of the many contributions of John Dewey has, to some extent, obscured his very real gift to educational psychology. His psychological contributions, since they are imbedded in the matrix of his other interests, sometimes are not recognized for what they are. Crissman (1942) for example, in order to formulate a résumé of the psychology of Dewey, had to search most of his most important works ostensibly concerned with other problems than those of psychology. The fact that Dewey performed little or nothing in the way of data research also strengthened the tendency to minimize his furtherance of educational psychology.

It was during the period 1894–1904, when head of the Department of Philosophy and Education of the University of Chicago, that he performed his more distinctively psychological contributions. Later, when he was Professor of Philosophy at Columbia University, his interests were exclusively in the philosophical and social spheres.

Although preceded by James, Dewey helped to found functionalism as a school of psychology in the late '90's. Along with Tufts, Mead, Moore, and Angell, there came into being the Chicago school of philosophical pragmatists and psychological functionalists. (Dewey, himself, preferred the term "instrumentalism" over pragmatism.) Dewey criticized both the "idealistic" views of faculty psychology and the "realistic" views of a purely biological psychology for their neglect of the social facets of human nature. Both of these contrasting views saw the individual child in isolation. He insisted that learning and education were social in character. The child was conceived by him to be active in the process of growth, and not the passive recipient of environmental influences. In this connection it is important to note contrary to a widely held opinion that he stressed not only the conditions which make for adaptation of the environment to the individual child, but also the adapting of the child to that environment.

Since Dewey's psychological views are perhaps not as well known as those of some of the other pioneers a short description is appropriate. Dewey's fullest expression of functional psychology came in his article, "The Reflex Arc Concept in Psychology." He wanted it to be seen that, just as the stimulus calls forth the response, so too does the response bear on the stimulus, and that the whole reflex arc, instead of being a closed unit,

is but a link in a chain of preceding and succeeding arcs. This chain gives us an effective instrument for effecting successful coordinations which aid the organism in its efforts toward the attainment of a goal. This is an adaptive process. The basic unit of Dewey's psychology was habit. Habits are acquired dispositions which are dynamic and persistent. They form the basis of impulses, emotions, motives, desires, perceptions, imagination, thought, meaning, object, mind, consciousness, and self. Many of Dewey's contributions had been to the area between the philosophy of education and psychology in which he would draw upon both fields. In *How We Think* he discussed the use of intelligence and an intellectual method assimilated to the behavior of the individual as he solves his problems. The individual needs to acquire those habits through which his thinking comes under his control. His thought becomes an instrument of experience.

Dewey's most closely related educational contributions deserve mention. He was the intellectual father of progressive education, with his emphasis upon personal interests, social factors, and practical activities. The appearance of differentiated curricula, the activity program, elective subjects, learning readiness programs, and an increased concern with the individual student, all attest to the changing views of educational functions in which progressive education was greatly interested. Although Dewey was important in these developments, he was not alone, nor was he guilty of promulgating what later came to be regarded as the excesses of that movement. Progressive education had its roots in psychology. As Krugman (1948) reminds us, progressive education consists essentially of the application of mental hygiene to education.

It was through the work of these pioneers that educational psychology was founded. Although working upon different problems and using diverse approaches, they had in common a desire for objectivity and a conviction that through quantitative research and measurement, educational psychology could be placed upon a firm footing. After these pioneers, and often as their students, came the first and subsequent generations of educational psychologists.

Closely related developments

Of necessity, this account of events related to the development of educational psychology occurring during these years must be very condensed. Attention will be confined to the development of, and the need for, courses in educational psychology; the beginning of the university study of education; the appearance of graduate schools; educational research through the measurement of achievement; and the beginnings of the psychology of learning.

Turning to courses in educational psychology, although it was not the first public normal school, the school at Oswego, N.Y., established in 1863, is important in that from the onset it had a course entitled "Child Study." In the years that followed the normal schools or departments of education being established routinely included courses in child study or educational psychology. The first of these terms was popular-

ized by Hall in the nineties, but after publication of Thorndike's *Educational Psychology* the latter term was the one preferred.[i]

In educational circles in the late 1880's the need for courses in educational psychology was beginning to be voiced explicitly by educators. In 1888 in San Francisco the meeting of the National Educational Association devoted considerable attention to educational psychology and its role in the training of teachers. Blair (1948) tells us that Parr, for example, after speaking of the fact that, all too often, educational psychology is taken to mean merely the study of general psychology with stray observations about children, went on to indicate that eventually it will be the subject arising from applying the principles of general mental science to the conscious process of development. He indicates that at the same meeting, Baldwin stressed the value of educational psychology for placing the teacher in a position where he can better understand the pupil, and hence, more intelligently lead him. Similarly, Hodgin considered educational psychology imperatively necessary.

Turning to the university study of education, the first permanent chair in education, a Professorship of Philosophy and Education, was established at the State University of Iowa in 1873. Up to 1890 less than a dozen chairs in education had been established in the United States. The work of these professors was largely limited to historical and philosophical studies. Thereafter, the number began to increase rapidly and the nature of their research to change in the direction of the collection of quantitative data.

What was apparently the first department of education, the Department of the Science and Art of Teaching, was founded at the University of Michigan in 1879. This was followed by the establishment of a Department of Pedagogy at the University of Washington in 1881 and similar departments at the University of North Carolina and at the Johns Hopkins University in 1884. What was to become of the foremost university training schools for teachers in the United States, Teachers College of Columbia University, was founded in 1888. Another leading school of education, that of the University of Chicago, was organized in 1900. In 1896–1897 the United States Commissioner of Education reported that 220 out of 432 institutions of higher learning offered courses in pedagogy.

Meanwhile, graduate schools in the modern sense were in the process of development. Before 1850 in the United States graduate education was a sporadic, unorganized affair. The graduate schools of Harvard and Yale, already in existence, were very much overshadowed by their powerful undergraduate colleges to which they were subordinated. The first earned Ph.D. degree in the United States was conferred by Yale University in 1861. Johns Hopkins opened its doors as a graduate school in 1876. For the first time the ideal of research as central, borrowed from the Germans, was given full and acknowledged impetus. A graduate school was established at Columbia University in 1880. Clark University, established in

[i] The first publication to bear the title *Educational Psychology,* otherwise unimportant, was a booklet published in 1886 by Louisa Hopkins (Roback, 1952.

1889, under the dominating leadership of G. Stanley Hall, soon thereafter offered thorough graduate training in a limited number of fields. In 1892, two years after its establishment, the University of Chicago embarked actively on its career as a leader in university instruction. Shortly before the beginning of the present century three state universities joined the group offering thoroughgoing graduate instruction—the University of Wisconsin in 1892, the University of Nebraska in 1895, and the University of Kansas in 1896.

Educational research, as differentiated from the narrower field of research in educational psychology, had its beginnings before the present century. Only one event in the broad history of educational research is discussed. This is the beginning of the measurement of achievement. Objective achievement measurement in education had its beginning in 1897 with the publication of two papers on achievement in spelling written by Joseph M. Rice (1897), a physician turned educator. The question to which he addressed himself in these papers was whether or not schools spending varying amounts of time in spelling drill showed corresponding difference in skill of their pupils. From his results he concluded that since there was no consistent relation between the achievement exhibited and time spent, the differences he found were due to differences in the quality of teaching. By modern standards the design he used in the studies was inadequate, but his critics were no better equipped than he in terms of sophistication of methodology. A country-wide controversy sprung up about his criticism of the quality of teaching which lasted for some years thereafter.

Another development related to educational psychology, taking place during these years, was the study of the psychology of learning. A short account of its beginnings is appropriate. It was the work of Ebbinghaus, first published in 1885, which had forcefully called the attention of psychologists to the possibility of controlled research in learning. His work demonstrated that learning could be measured by accurate techniques. Thereafter, research in this area was pursued with enthusiasm. For example, before 1900 Meumann began research in experimental pedagogy and wrote an influential book which was subsequently translated into English. Despite the fact that the conceptual significance of learning was not fully recognized, as shown, for example, by the failure of Baldwin's *Dictionary of Psychology,* published in 1902, to even list learning as a topic, research activity was not inconsiderable before the new century. In a selected bibliography, published nearly thirty years ago, McGeoch (1933) cited 1200 titles. Six of these papers were published before 1890 and 60 more articles before 1900. Before 1900, investigators had performed studies of the influence on learning and retention of age differences, interference, the curve of learning, reflex "control," motor skills, direction of association, the influence of distraction, the distribution of practice, diurnal variations, fatigue, frequency, habit formation, individual cues, intelligence, knowledge of results, degree of meaningful organization, memory span, racial differences, recitation, sense modality, time interval in associations, transfer, varia-

bility and difficulty of material, the whole-part problem, affective memory, childhood memories, context conditions, curve of retention, incidental memory, recognition, reminiscence, and repeated reproduction. Many of these studies were by researchers who did not systematically follow through after their first effort in this area. Content to publish one or two studies, thereafter no sign of continued publication could be observed. Nevertheless, a respectable beginning on problems of learning had been made before the turn of the century.

The Incubation of Educational Psychology: 1900–1918

If the beginnings of educational psychology took place before 1900, most of the next two decades was its period of incubation. During these years psychologists began to specialize in educational psychology. It was the period of Edward L. Thorndike and Charles H. Judd, the first educational psychologists. It was also a time of intense activity concerning the learning and reading processes, activity in large measure stimulated by these two men. Intelligence and achievement testing also received considerable attention during these years.

Edward L. Thorndike

The first man to deserve to be called an educational psychologist in the modern sense of the term is Edward L. Thorndike. Almost his entire academic life was spent at Teachers College, Columbia University, where for over 40 years he labored on problems of educational psychology. Generations of students, including both prospective teachers and psychologists, came under his influence. He held that educational procedures of all sorts should be based upon results of psychological research, not upon opinions. He stated his position bluntly and succinctly in the preface to a book published three years after the turn of the century. "This book attempts to apply to a number of educational problems the methods of exact science. I have therefore paid no attention to speculative opinions and very little attention to the conclusions of students who present data in so rough and incomplete a form that accurate quantitative treatment is impossible" (1903, p. V).

It was Thorndike who patiently and consistently first systematized the study of learning. Beginning his work in animal learning while a graduate student at Harvard under James, Thorndike performed quantitative experiments on animal learning. The 1897 study of the learning of cats in a puzzle box is a landmark in the history of animal psychology. Although a little later he concentrated on learning in humans, he maintained the same interest shown in this study in that he continued to be concerned with the nature of the learning curve and the conditions which affected it. His early research at Columbia eventuated in his *Elements of Psychology* (1905) which contained his formulation of a variety of "laws of learning"—exercise, effect, and readiness.

Thorndike and Woodworth at the turn of the century published their famous paper on transfer of training. Thorndike followed this with a series of studies

designed to verify or to refute the doctrines of transfer of training and formal discipline. The general conclusion from these studies was that ideas or habits acquired in one sphere of activity were transferred to another sphere only if there were common elements shared by both spheres.

He was also interested in arriving at some conception of the nature of learning. Learning was seen as forming bonds or connections between stimuli and responses. Hence he combatted the faculty point of view that learning was the training of the potential faculties. Some years later, with the appearance of studies by others on the basis of which insight was postulated as a form of learning, he vigorously disagreed, stressing that the formation of bonds of association was a gradual process even though they sometimes gave the impression of sudden appearance. When in 1913–1914 his three volume edition of *Educational Psychology* came out, most of the experimental material included was derived from his own original investigations. Although his theoretical interpretation of his studies of learning has since been questioned, no one can deny his great influence upon the field of learning.

Although at the core of Thorndike's efforts were his numerous studies of learning, it is astonishing to find in how many other areas he worked in these early years of the century. In 1902 he published the first work to make generally accessible the statistical approach to the problems of education, sociology, economics and anthropology, *The Theory of Mental and Social Measurements.* In it he considered measures of central tendency, variability, and correlation and so made available to the student an understanding of the treatment of data which was beginning to flow in increasing amounts from the laboratory and classroom. In 1907 he prepared a report for the U.S. Bureau of Education on the elimination of children from the schools who could not profit from the curriculum, a forerunner of many other reports on this and related problems. Thorndike promoted the measurement of ability in school children and developed scales for the measurement of subject-matter fields. It was in 1908 that Stone published his arithmetic test, which is considered to be the first standardized achievement test. Before 1917 Thorndike and his students or associates constructed and standardized a considerable number of other tests. By the name of the investigator, the subject matter field and the date of appearance they were as follows: Curtis, Arithmetic, 1908; Thorndike, Handwriting, 1909; Hillegas, English Composition, 1912; Buckingham, Spelling, 1913; and Thorndike, Reading, 1914, 1916.

Charles H. Judd

Sharing with Thorndike the honor of being a pioneer educational psychologist was Charles H. Judd. He was trained at Leipzig by Wundt, receiving his degree in 1896. His *Genetic Psychology for Teachers* exerted considerable influence upon school practice. In it he stressed the psychology of the school subjects of reading, writing and arithmetic and forcefully stated the doctrine of biological and psychological development. After some years of other academic ap-

pointments, he came to the University of Chicago as Director of the School of Education, a post he held from 1909 to 1938. He took part in many educational surveys and edited various monographs and periodicals, including the *Elementary School Journal* and the *School Review.*

It was in the field of the psychology of reading that Judd's work was particularly outstanding. While at Yale University, he developed a method for the photographic study of eye movements. The studies of reading which were carried out at the University of Chicago were largely under his inspiration or direction.

There had, of course, been earlier work on reading. Perceptual studies of the process of reading had been conducted as early as 1844 by Valentius. This was followed by the work of Cattell, Erdmann and Dodge in such problems as the study of eye movements.

Gray (1960) has made a tabulation of the number of scientific studies relating to reading published in the United States and England by decades beginning with 1881. In the three decades before 1911 there were 32 such studies. But in the decade 1911–1920 alone there were six times as many studies, with 200 reported by him. According to the same reviewer prior to 1910 most of the research had been concerned with the psychology and physiology of reading. The period from 1911 to 1920 is characterized by him as a transitional one since it marked the beginning of an interest in a broader conception of the field and showed a more clear recognition of the need for applying objective techniques to classroom problems. Moreover, it was during this period that the new tests were introduced, making it possible to study under classroom conditions large groups of children.

Intelligence testing

Utilization of the Binet test in the United States came about particularly through the efforts of H. H. Goddard, F. Kuhlmann, and L. M. Terman, three Ph.D.'s from Clark University. In 1908 Goddard translated the Binet Scale into English and applied it to American children. In 1912 Kuhlmann introduced in the United States a revision of the Binet Test. In 1916 Terman extended and modified the Binet Scale, retaining its good features and eliminating some of its weaknesses, so that it was especially adapted for use in the United States. Although the first two aforementioned forms were used in research, it was Terman's revision, of the so-called Stanford-Binet, that became the indispensible, widely used instrument for testing the intelligence of individual children. To designate the degree of intelligence of the subject, irrespective of age, in 1912 Stern had advanced the concept of the intelligence quotient. This useful procedure was adopted by Terman in his revision of the test.

By 1917 the literature on the Binet tests had proliferated to the extent that in that year Boardman was able to report a bibliography of 344 articles and books. Of the writers, H. H. Goddard was most prolific, having published 28 articles concerning the Binet between 1908 and 1916. Binet and Simon, either alone or in collaboration, supplied 19 titles. J. E. W. Wallin published 16 reports. The pub-

lications of L. M. Terman, by this time, numbered 15, including one published in 1906 concerned with a specific test from the scale. His first report on the Binet-Simon Scale as a whole was published in 1911.

Achievement testing

Prophetic of the many later studies of learning in the classroom was the one by Ayres, who studied individual differences in learning through differences in school achievement. As a means of presenting his results he used the method of grade overlapping. His results received wide publicity.

The ten years after Rice's 1897 publication was the period in which the measurement idea incubated. The work of Thorndike, already mentioned, was important during this decade. In 1912 achievement tests were used for the first time in a large school survey, taking place in New York City. Over 125 similar surveys were conducted in the ten years that followed. Despite this progress, objections to the use of standardized achievement tests were numerous. A favorite device of protagonists for the new tests was the study of the reliability, or, rather, the unreliability, of essay examinations which were the long-entrenched alternative to achievement tests. The studies of Starch and Elliott uncovered what was taken to be an amazing lack of agreement among teachers in grading essay-type examinations in the various high school subjects. This and similar evidence resulted in an increasingly favorable reaction to objective tests.

By 1915 enough support for objective testing was available at a meeting of the National Council of Education to bring about a clear-cut decision in its favor. This made such testing respectable in general educational circles. Objectivity and measurement were now established as generally accepted aspects of educational research. It was no accident that it was during these years that Thorndike offered his dictum that everything exists in some amount and that this allowed measurement of that amount.

Other developments

The National Society for the Study of Education was founded in 1902, but had derived its being from the National Herbart Society founded in 1895. Yearbooks published by the Society include many directly pertinent to educational psychology. They continue to be extremely influential.

The educational research bureau was one of the characteristic developments in the second decade of this century. The first such bureaus were established in connection with city school systems —Baltimore in 1912 and others less than two years later in Rochester, New Orleans, New York City and six other cities. College and university research bureaus were established in 1913 and 1914 at Oklahoma University, Indiana University, the State University of Iowa, and what is now Kansas State Teachers College. By 1925 there were 22 college educational research bureaus, and by 1932 there were 64. State department research bureaus were also organized just before and after World War I. By 1915 there were enough of these bu-

reaus for the National Association of Directors of Educational Research (now the American Educational Research Association) to be formed. The *Journal of Educational Psychology* was founded in 1910.

Growth of the field

Some idea of the growth of educational psychology over the years to about 1918 may be gained from examination of the relative percentage of publications in this field as compared to those found in other areas of education. Franke & Davis (1931) made a survey of published educational research, irrespective of field, over the four decades of 1890, 1900, 1910, and 1920. They classified into fields the articles appearing in 13 educational periodicals (*Education, Educational Review, Journal of Genetic Psychology, Teachers College Record,* etc.). In 1890–1899 child psychology had the highest percentage of articles devoted to it, with curriculum, character, educational measurement and statistics coming next in that order. Next to the last, with about 3 per cent of the publications for that decade, was educational psychology (so defined as to refer mostly to studies of learning). The decade 1900–1909 showed similar percentages. By 1910–1919 educational psychology had moved up to be tied for third place with about 10 percent of the pages of these journals devoted to it. This position was maintained in the following decennial period, 1920–1929. By 1918 educational psychology had been established as an important field of research and its period of incubation drew to a close.

The Recent Past in Educational Psychology: 1918–1941

A survey of 10 years of educational research covering the period of 1918 to 1927 was published by Walter S. Monroe (1928). His choice of 1918 as beginning a new period of educational research was based on the following considerations: It was during 1917 that psychological testing of Army recruits began, which tended to create a popular interest in the measurement of intelligence; it was in 1917 that the Iowa Child Welfare Research Station was authorized; and it was in 1919 that the Commonwealth Fund, noted for its subventions for educational research, was established, as were the American Council on Education, and the Bureau of Educational Research at the University of Illinois. I would add that it was in 1917 that Ayres stated that measurement had been accepted by the public, and it was in 1919 that, as Buckingham reminded us, test materials were first produced by commercial publishers. Gates and his associates (1948) assert that about 1920 educational psychology could be claimed to have taken definite form. Moreover, a contrast in the treatment of psychology in relation to education became apparent at the beginning of the twenties. This may be illustrated by two widely read authoritative books on the history of education published in 1905 and 1920. One book by Paul Monroe (1905) closes his discussion of psychology with the work of Herbart. The other by Ellwood Cubberly called psychology "the master science" (1920, p. 255), and proceeded to document this claim. For these reasons the

recent past in educational psychology is considered to have begun about 1918. Events from the time of World War II onward are too close to hope to see them in historical perspective. Hence, events after about 1941 are not considered.

For the recent past it is impossible to approach the history of educational psychology biographically within the limitations of article length. Too many men have contributed to this period. Even though some of them are mentioned in the account to follow, perhaps an equal or even greater number of other important contributors have been neglected.

Major researchers of the earlier years

In their survey Monroe and his associates (1928) were able to assemble about 3700 "worthwhile" references to educational research for the decade 1918–1927, the earlier years of the recent past. From among these they selected five major research projects as the most highly important that were published during that period. The first of these was "the Chicago reading studies." Between 1920 and 1925, Judd and his colleagues, Buswell, Terry, and W. S. Gray, published reports of six major studies carried out in the reading laboratory at the University of Chicago. The studies concerned the eye-voice span, the reading of numerals, the development of reading habits, remedial cases in reading, the kinds of silent reading, and, finally, a summary of investigations related to reading. The Thorndike study of the measurement of intelligence in the period 1922–1925 which culminated in the book by Thorndike and his associates, *The Measurement of Intelligence,* was another project selected by Monroe. The studies of genius conducted by Terman and his co-workers, including Katharine Cox, were also selected as among the important studies of that decade. This project was begun in 1921 and the decade saw publication of both Volume I and Volume II of the study. The fourth major project was the nature-nurture studies reported in the *Twenty-seventh Yearbook of the National Society for the Study of Education.* The yearbook was devoted to a large number of papers concerned with the influence of nature and nurture upon intelligence and achievement. The studies dealt with the causes and significance of the large individual differences in intelligence and achievement test results. Among the participants were Barbara Burks, Truman Kelley, E. L. Thorndike, H. E. Jones, F. N. Freeman, K. J. Holzinger, Florence Goodenough, Gertrude Hildreth, Arnold Gesell, Leta Hollingworth, W. McCall and J. Peterson. The papers included one by Burks and Kelley on statistical hazards in nature-nurture study; several on family resemblance in intelligence test scores; others on the relation of intelligence to social environment, to race differences, to schooling, to health, to physique, and to coaching. Still others were devoted to the relation of achievement to intelligence, school attendance, school methods, school expenditures, effort, and mechanical ability. Since potentially any kind of educational research might have been selected, it is significant that it was only the fifth study which did not fall

within the scope of educational psychology. This was an investigation, begun in 1921, concerned with educational finance.

Intelligence testing

As a field of study and application educational psychology received considerable stimulation during the earlier years of the recent past from the development of the intelligence test arising from the work of Binet and those that came after him. Several lines of evidence attest to the attention to intelligence testing. Although it was opposed at first, acceptance on the part of educators of the concept of intelligence as the mass indicator of intellectual maturation in their students was finally won. Parents, by and large, came to accept application of such tests although perhaps grudgingly. With changing views of the function of the schools, no longer was such measurement seen as an invasion of privacy as it was in the days of Hall when he wished to get height and weight measurement of pupils in the schools. The amount of research appearing before 1917, which was stimulated by the Binet Test, documents the interest. Even more numerous studies appeared after World War I. The development of the sub-specialty of the mental tester and, indirectly, of the clinical psychologist, was fostered by this interest, since administering the Binet became what was practically a full-time occupation for a considerable number of persons.

The Stanford-Binet was seized upon by educators and psychologists when it was recognized that it met certain already existing needs. Compulsory school attendance, the increase in the length of school period, and the increasing number of backward students in the schools created a favorable situation for its avid use. These same educational problems also made desirable the development and utilization of instruments for the large scale assessment of groups.

Group testing

World War I gave special impetus to the interest in group intelligence testing. There had been some preliminary work by Binet and Simon, and selection and the use of group intelligence tests in the U.S. had been begun by Arthur Otis and L. L. Thurstone a few years before the entry of the U.S. in the War. For example, in 1915 the latter had started a group testing program for prospective students at Carnegie Institute of Technology. It was the former, however, whose first work with group tests helped in the development of the Army Alpha and Army Beta which were used in selecting Army recruits. About 1,750,000 men were given these tests.

The twenties saw the application of group intelligence tests to school problems. The dates that representative well-known tests were published is of some interest. *The Otis Group Intelligence Scale* (*Advanced Examination*) appeared in 1918. *The Terman Group Test of Mental Ability* and *The Haggerty Intelligence Examination, Delta II* both appeared in 1920. *The Otis Self-Administering Tests of Mental Ability* were published in 1922, and the *American Council Psychological Examination* of

the Thurstones first appeared in the 1924 edition. Both the *CAVD Intelligence Scale* of Thorndike and the *Kuhlmann-Anderson Intelligence Tests* appeared in 1927. Group testing of intelligence was established as an important phase of educational psychology by the beginning of the thirties.

Although attention to theoretical formulations was not entirely absent, most psychologists were relatively uninterested in the nature of what they were so busily measuring. In an attempt to remedy this situation, a symposium was held on the meaning of intelligence and published in the *Journal of Educational Psychology.* The reactions of the participants were diverse and contradictory but the overall effect of the symposium was salutary. S. L. Pressey probably represented the majority of psychologists of that time when he asserted at the symposium that he was more interested in what a test would do than he was in defining intelligence. Nevertheless, it is of some interest to examine briefly some of the participants' remarks. The points stressed were as follows: Thorndike—the goodness of response from the standpoint of truth; Terman—abstract thinking; Calvin—learning to adjust; Pintner—a means of adapting to new situations; V. A. C. Henmon—a combination of intellect and knowledge; Peterson—a mechanism for adjustment; Thurstone—a combination of inhibitive capacity, analytical capacity, and perseverence; Woodrow—the capacity to acquire capacity for performing acts successfully; Dearborn—the capacity to profit by experience; and Haggerty—the activity of a group of complex mental processes. The functional cast of the definitions by these leading educational psychologists is overwhelmingly evident.

Attempts at the measurement of personality lagged behind those concerned with the measurement of intelligence and achievement. The prototype of paper and pencil personality questionnaires was the Woodworth *Personal Data Sheet,* begun during World War I, not used during the war but appearing shortly thereafter. The *Mental Hygiene Inventory* of House appeared in 1926. It was not until 1929 that the *Personality Schedule* of L. L. and T. G. Thurstone was published. In 1931 both the *Personality Inventory* of R. G. Bernreuter and the *Emotional Maturity* (*E.M.*) *Scale* of R. R. Willoughby made their appearance.

Another form of psychological testing important to educational psychology, that of the testing of interests, came into prominence in the 1920's and the 1930's. Earlier Herbart, Dewey, Hall and Thorndike, each in his own way, had emphasized the importance of the factor of interest to the psychology of education. As a consequence, research on the measurement of interest was inevitable. Work was begun at the Carnegie Institute of Technology in 1919 upon the earliest standardized interest inventory in the sense that term is used today, *i.e.,* statistical evaluation of the instrument and the working out of an objective scoring method. Clarence Yoakum, Bruce Moore and J. B. Miner were important figures in this early work. Cowdery revised the Carnegie Inventory in 1924 and used it for differentiating groups on the basis of their interests. E. K. Strong took up this form of investi-

gation. The now very well known Strong *Vocational Interest Blank,* which first appeared in 1928, was a thorough revision and extension of the Cowdery Inventory. Many studies using this measure as a research instrument appeared in the thirties.

Under the leadership of Ben Wood, The Cooperative Test Service began in 1932 publishing a considerable number of comparable forms of tests.

Child development

Hall and others had already established an interest in the growth and development concept. Even before the beginning of the period in question a considerable amount of material was available. One illustration will suffice. In March, 1914, a bibliography of the available representative books in child study was published by Louis N. Wilson, then librarian of Clark University. The first section was devoted to general child psychology and, aside from the usual classic sources of Darwin, Hall and Preyer, 12 other books published since the turn of the century were included. Among the authors cited were Edwin A. Kirkpatrick, Kathleen C. Moore, Milicent Shinn and Edward L. Thorndike. Eleven books or monographs concerning the application of child study to education were listed. The health of the child received eight citations including a book by L. M. Terman, *The Hygiene of the School Child.* Four references to defective children and a similar number of reminiscent studies of childhood followed. The bibliography concluded with mention of the four child study journals then available: *The Child, Child-Study, Journal of Educational Psychology,* and the *Pedagogical Seminary.*

Arnold Gesell, one of Hall's protégés, eventually combined pediatric and psychological interests in his meticulous patient studies of mental and physical development in children. After taking his degree in psychology, he had studied medicine in order to learn more about the physical bases and psychological processes of growth. In 1911 while connected with Yale University, he was assigned space in the New Haven Dispensary for the establishment of a psycho-clinic for children. This was the beginning of the Yale Clinic of Child Development. In 1924 came his first contact with the cinema when he prepared a film on the behavior development of the pre-school child from early infancy to school entrance. One of his first major works, *The Mental Growth of the Pre-School Child,* appeared in 1925. This volume was one of a long series devoted to what are now his well known studies of child development. About 1926, the Clinic started to take on its modern form, complete with one way vision screens, research associates and a systematic plan for the recording of infant behavior.

One of the first systematic studies of physical and mental growth in a variety of its manifestations was carried on under B. T. Baldwin, director of the Iowa Child Welfare Station from 1917 to 1928. In 1921 Baldwin published a classic study of the physical growth of children with primary emphasis on anthropometric measures. He worked out the growth curves of these measures including some for individual children, repeatedly measured.

In 1922 W. F. Dearborn started the so-called Harvard Growth Study in which he followed a population of children from the time they entered the first grade until they finished high school. Both mental and physical tests were used.

The technique of longitudinal study was followed at various other so-called institutes of child welfare which began to function throughout the country during these years. Funds from Laura Spellman Rockefeller and the General Education Board supported this development. In addition to those already mentioned, institutes at the University of Minnesota, the University of California, the Merrill-Palmer School and the Fels Research Institute became active, especially in areas where long continued studies of children of wide scope were necessary. In 1935 the Society for Research in Child Development was organized.

The influence of schools of psychology

Behaviorism, psychoanalysis and the Gestalt psychologies had appeared on the psychological scene before the beginning of the period under discussion. However, some time elapsed before their impact upon educational psychology was perceptible. The onset of their influences occurred in the twenties and thirties.

The militant espousal of an objective approach, the behaviorism of John B. Watson, had made its appearance beginning in 1912, culminating in his 1919 publication, *Psychology from the Standpoint of a Behaviorist.* It took some years for its effect to be seen in educational circles.

What is the nature and extent of the effect of behaviorism upon educational psychology? Butts and Cremin (1953) consider behaviorism to have performed the function of serving as a more extreme version of associationism. They do not go on to cite specific instances of its effect upon education. In fact, here or elsewhere, specific influences of behaviorism upon educational psychology are not too evident. An illustration drawn from a review of behaviorism is relevant. In 1938 Harrell and Harrison (1938) published a paper entitled, "The Rise and Fall of Behaviorism." They cite 426 references. These references were searched for items of educational import. They mention incidentally that J. S. Gray (1935) wrote a manual of educational psychology from the standpoint of a strict objectivist. This is their only reference to educational psychology. Turning to this manual, in the preface Gray made it clear that he adheres to no specific school, and that he wished to stress objective data, particularly what he called "biological data," leading him to stress such work as that of C. J. Herrick, G. E. Coghill, and C. M. Child. No evidence is available to the writer about the reception of the book. Others similar in spirit and addressed to problems in educational psychology do not seem to have been published. To return to Harrell and Harrison, search of their bibliography reveals one reference in the *Journal of Educational Psychology,* one in the *Hibbert Journal* and one in *School and Society.* Searching their contents showed that one had but an

incidental mention and both of the others were critical of behaviorism. Of course, I am confident that if a patient, complete search was made of educational periodicals other articles referring to behaviorism would undoubtedly be found. However, no article of this sort seemed to have reached the stature of being considered as a major contribution to educational psychology so far as the writer is aware. One is left with the impression that behaviorism during these years (or later) did not have much specific effect upon educational psychology.

The Gestalt psychologies arising from the work of Wertheimer, Köhler, Koffka, and Lewin came into educational psychology in a variety of ways. The publication of the book by R. M. Ogden (1926), *Psychology in Education,* called attention to the value of Gestalt to students of educational psychology. In 1931 Seltsam was stimulated to report the main contentions of Gestalt psychology in the *Journal of Educational Psychology,* because of what he said was an increasing interest and concern with this approach for which he cited specific instances. The effect of Gestalt psychology upon education has helped to give an integrationist view of human behavior. The influence of Gestalt psychology was probably enhanced by its affinity with the independently developed progressive education movement. Gestalt psychology and progressive education have many points of compatibility, if not identity. He suggested as illustration, the Dewey-Kirkpatrick principles of total rather than isolated learning, and their premise that parts of learning should be learned in relationships to the whole to which they belong. A number of parallels are dealt with in detail. Not only has Gestalt psychology integrated to an extent with progressive education, but it has also served a function, in the view of many workers, as a corrective to the extremes of connectionism and behaviorism. The influence of the work of Lewin upon social psychology and, indirectly, upon social psychology of childhood is self-evident.

It is probable that no one would deny that psychoanalysis had an influence upon educational psychology. As was the case with behaviorism, it is difficult to be specific about this matter. The problem is complicated by the fact that its influence has probably been indirect in that some other vehicle such as child guidance clinics, the progressive education movement, mental hygiene and the like, were first influenced, and they in turn influenced educational psychology. That it has had an influence is attested to by the study performed by Park (1931). She analyzed the contents of general psychology texts and found that not a single one of the books examined published between 1917 and 1930 failed to show, either explicitly or implicitly, the influence of psychoanalytic doctrines. She concluded that this influence increased steadily between 1910 and 1930. It is plausible to believe that psychoanalysis had an influence upon guiding the thinking of educational psychologists in the direction of emphasis on the importance of the early years, stress on the irrational facets of the child's makeup, on the necessity of warmth and permissiveness and, more

generally, on increased attention to all facets of the personality and makeup of the child rather than to his mind or intellect alone.

Courses in educational psychology

Using 1927–1930 as the base years, Gates (1932) analyzed various studies concerned with the programs of courses in educational psychology in representative teacher training institutions. By and large, these studies had all found a considerable increase in the total number of courses over those offered during the previous decade. As Gates indicates, the expansion consisted mainly of developing as separate courses some of the topics formerly treated more briefly in the general courses in educational psychology. He also attributed the increase to the fact that there has been a lengthening of teacher training with the older types of two-year normal schools increasing their offerings to three or four years. Meanwhile, the liberal arts colleges had been adding to their curricula a sufficient number of courses in education to meet the requirements for a teachers certificate. Up to this time this had been a relatively infrequent practice.

The concept of evaluation

It was during the thirties that Ralph W. Tyler of Ohio State University and his associates gave "a new direction" to measurement through the concept of evaluation. The term refers to measurement of student progress, not so much in terms of achievement as the term is ordinarily used, but rather in terms of the broader social objectives of education. This means that a value scale is substituted for a measurement scale. Measurement is a matter of "moreness." Value has to do with appropriateness, which falls on an optimum point along a scale with the rest of the scale extending on both sides of this point. In this perspective, amounts beyond this optimum may be as damaging as having too little. The nature of evaluation and some of the cautions in its conceptual application are given in Lorge (1941).

Progressive education

In a report published in 1941, a committee of the Progressive Education Association reported an analysis and summary of recent comparative research studies of old and new methods in education. The committee that wrote this report was chaired by G. Durwood Baker, a superintendent of schools, and had as some of its members R. Travers, A. Eurich, I. Lorge, P. J. Rulon, R. Tyler, Elizabeth Woods and J. W. Wrightstone. In the preface they state that it was only during the '30's that public school systems throughout the country attacked the problem of revising and modernizing their courses of study. This volume is an important summarization of what has been referred to as progressive education. They define progressivism as carrying out a great deal of the learning going on in the school through activities at the expense of a smaller amount of time spent in routine drill. They also emphasized the importance of training for a living, learning by doing, and educating for participation in society. All these are to be

carried through by an integrated curriculum, *i.e.,* drawing upon any field of knowledge that will assist the child in fulfilling these objectives.

After reporting a variety of surveys and studies comparing and contrasting older and newer methods of teaching they reached the conclusions that the newer methods do not result in the loss of academic proficiency in the usual school subjects and that, invariably, there is a definite gain in such characteristics as initiative, skill in dealing with problems, and knowledge of world affairs and social participation. They also point out, as an implication of the evidence, that each child has an individuality of his own and that learning in the school must be continuous and not a thing apart.

School achievement

W. S. Monroe (1945) in contrasting the measurement of school achievement of 1945 with that of 1920 points to the sheer increase in number and kinds of tests, but stresses the ingenuity shown in devising new types of objective exercises, greater attention to the compatability of the exercises and the abilities one is attempting to measure, and greater attention to and complexity of item analysis.

Reading

According to Gray (1960), since 1920 the scope of studies of reading has broadened considerably. Since that time recognition has come that this field includes problems that arise before the children begin to actually read and that studies of adult readers are also necessary and relevant. Increased appreciation of the complexity of the problem of reading is also apparent.

Learning

In 1940 Stroud published a careful selective review of studies of learning in the school situation. Despite this selectivity he referred to 256 titles. He closed the review with general evaluative statements which are of interest. He concluded: (1) that the discrepancies between the laboratory and school setting are not a serious obstacle to interpretation; and (2) that the research in the school setting compares not too unfavorably with that done in the laboratory, although not as good as the very best done in the laboratory. He goes on to a plea for closer cooperation between experimental and educational psychologists. Munn (1942) reached a somewhat different conclusion after reviewing the classroom application of learning. Although there are leads, he feels that in most instances application to classroom conditions and curricular materials is yet to be worked out. He sketched the two extremes of the positions taken by various educational psychologists. On one hand, there are those who would transfer practically any kind of results of the study of learning to the classroom; at the other, there are those who consider relevant only those studies performed in the classroom itself. He feels the truth lies between the extremes.

The 1942 or Forty-first *Yearbook of the National Society for the Study of Education* is a landmark in the history of educational psychology. It was de-

voted to the psychology of learning. The first of the two sections was concerned with theories of learning. Guthrie, Hull and Lewin presented their own views of learning while connectionism was represented by articles by Gates and by Sandiford, and field theory by Hartmann as well as by Lewin. An attempt at a reconciliation of the learning theories by McConnell closed this section of the *Yearbook.* He finds in their accounts points of synthesis and reconciliation. In fact, he lists nine points of similarity. He is commendably cautious in his interpretation, hoping that the consideration of the theories will stimulate further research. The second major section of the report was concerned with the implications for education of the various viewpoints expressed in the first section. Ryans discussed the implications for motivation; J. E. Anderson those for emotional behavior; Stroud those for practice; Horn those for language and meaning; Brownell those for problem solving; and Buswell those for the organization and reference to the curriculum.

Social psychology

It is prophetic that in 1941 Trow called attention to the fact that most educational psychologists had been content to limit themselves to the psychology of the individual, while neglecting the social facets of his behavior. The pressures for attention to social factors came not so much from educational psychologists as from so-called progressive educators, curriculum specialists, youth leaders and the like. Dewey was, of course, a prime force in this plea for the child to be seen as a social being.

The studies in social psychology relevant to education seem to have been published toward the end of the period of the recent past. To be sure, there had been earlier research descriptive of social development in the normative sense but little research on how this development took place or on the dynamics of the social forces impinging on the child. In the years between 1927 and 1939 Lehmann and Witty published on play activities, Maller gave his report on cooperation and competition, Goodenough published her extensive study of anger in young children, Isaacs' book on social development was published, the study of childhood conflicts by Jersild and Markey appeared, Lois Murphy published her book on social behavior and personality, Lewin, Lippitt, and White published a major paper on "social climates," and in the same year Dollard and his associates published their book on frustration and aggression. These developments in social psychology were important pioneer steps leading to the present day concern with the social aspects of educational psychology. Before these and similar studies made their appearance, educational psychology was primarily oriented toward increasing the efficiency of the learning of the formal subject matters of education, as the preceding discussion shows.

By 1941, the beginning of the extended present, this orientation is seen only as a point of departure since learning takes place everywhere and in all situations whether scholastic or not. The task of the school had broadened

beyond imparting formal knowledge and skills. The doctrine of the "whole child" brought in its wake concern over the social and personal development of the child. This state of affairs broadened the scope of educational psychology, but it also tended to blur the nature and extent of educational psychology. Guiding development in the schools had become increasingly complex.

Reflections on the History of Educational Psychology

The comments that close this brief survey of educational psychology are confined to some reflections on two of the major trends exhibited.

Another paper (Watson, to be published), concerned with the present status of educational psychology and of educational psychologists, is relevant in considering the significance of the facts that have been brought forth in this paper. In this other paper the nature of educational psychology is shown by presenting descriptions of its scope by various authorities, by examining available information about undergraduate course offerings, by comparing the content of textbooks in the field, by analyzing contributions to it from various other areas of psychology, and by discussing its reciprocal contributions to these areas of psychology. In the paper in question, there is also consideration of educational psychologists in terms of their number, their rate of growth, the positions they hold, the training they receive, the research they carry on, the values they hold, and the prestige afforded them by their colleagues. Thereafter, the various steps being taken to remedy some of the weaknesses that are recognized as existing in educational psychology and among educational psychologists are surveyed. It is significant that the major finding about the present status of educational psychology is widespread dissatisfaction expressed both by psychologists within the field and those outside of it. More specifically, educational psychology is found to have diminished in prestige; research in the field is felt to be uneven, and, on the whole, disappointing; feelings of estrangement are evidenced between psychologists in general and educational psychologists, and a field is noted that while broadening in scope is losing its former coordination. Can we see from this historical survey any of the roots of these sources of dissatisfaction? If one stands back from the mass of historical details that were presented and tries to see the broad outline, the reviewer suggests that two trends may be discerned.

First, there has been an overwhelming emphasis in educational psychology upon the practical application of psychology. Theoretical issues have been neglected. Unless it is argued this reviewer has been guilty of selective projection in writing this account so as to overstress the applied aspect and minimize the theoretical, this conclusion seems self-evident. There is a relative lack of theoretical orientation in educational psychology. Throughout the time span of the historical survey the research problems appear to be selected primarily because of pragmatic interests. Perhaps not since Dewey has there been a strong theoretical guideline for the conduct of research in

educational psychology. In a sense the connectionism of Thorndike was a militant espousal of the lack of necessity for a theoretical orientation. Behaviorism, psychoanalysis and Gestalt theory, although they made some impression on educational psychology, arose outside of the field and did nothing more than modify it in some respects. In no way can it be said that any or all of them did anything to revolutionize the field. The theoretical orientations that guide modern approaches to learning seem to have made astonishingly little impression upon research or application of the psychology of learning to the problems of educational psychology. They are not the foundation upon which educational psychologists build their work.

Second, the field of educational psychology has become more complex as the vision of what it encompasses has broadened. Originally concerned with learning and measurement, its scope has been extended with each succeeding generation, to the point where now the newest extension is social educational or educational social psychology. When one is reminded that this reviewer quite deliberately omitted consideration of the professional service functions, such as those in clinical and counseling activities which are also broadening the field, and confined himself to teaching and research, this sense of a broadened field is heightened. As educational psychology becomes broader, it continues to train its own specialists in whatever is the particular new area in question. Consequently, there is more and more separation of educational psychologists from the rest of their brethren. Most often trained in different schools, on different subject matter, and taking jobs in separate places of employment, they continue to lose contact with the rest of the psychological fraternity. At least in these respects the history of educational psychology shows that the field has not as yet realized its full potential.

References

Allport, G. W. The productive paradoxes of William James. *Psychol. Rev.,* 1943, **50**, 95–120.

Blair, G. M. *Educational Psychology, Its Development and Present Status.* Urbana, Ill.: Bureau of Research & Service, College of Education, 1948.

Boring, E. G. *A History of Experimental Psychology* (2nd ed.) New York: Appleton, Century, Crofts. 1950.

Butts, R. F., and Cremin, L. A. *A History of Education in American Culture.* New York: Holt, 1953.

Cowdery, K. M. Measurement of professional attitude differences between lawyers, physicians and engineers. *J. Personnel Res.,* 1926–1927, **5**, 131–141.

Crissman, P. The psychology of John Dewey. *Psychol. Rev.,* 1942, **49**, 441–462.

Cubberley, E. P. *History of Education.* Boston: Houghton Mifflin, 1920.

Franke, P. R., and Davis, R. A. Changing tendencies in educational research. *J. Educ. Res.,* 1931, **23**, 133–145.

Gates, A. I. The place of educational psychology in the curriculum for the education of teachers. In S. A. Curtis (Ed.), The direct contribution of educational psychology to teacher training. *Yearb. Nat. Soc. Coll. Teach. Educ.* 1932, **20**, pp. 21–35.

Gates, A. I., Jersild, A. T., McConnell, T. R., and Challman, R. C. *Educational Psychology.* New York: Macmillan, 1948.

Gray, J. S. *Psychological Foundation of Education.* New York: American Book, 1935.

Gray, W. S. Reading. In C. W. Harris (Ed.), *Encyclopedia of Educational Research.* (3rd. ed.) New York: Macmillan, 1960, pp. 1086–1088.

Harrell, W., and Harrison, R. The rise and fall of behaviorism. *J. Gen. Psychol.,* 1938, **18**, 367–421.

Krugman, M. Orthopsychiatry and education. In L. G. Lowrey & Victoria Sloane (Eds.), *Orthopsychiatry,* 1923–1948: *Retrospect and Prospect.* New York: American Orthopsychiatric Association. 1948, pp. 248–262.

Lorge, I. Evaluation: the new stress on measurement. *Teach. Coll. Rec., 1941,* **42**, 667–679.

McGeoch, J. A. The psychology of human learning: a bibliography. *Psychol. Bull.,* 1933, **30**, 1–62.

Monroe, P. *A Textbook in the History of Education.* New York: Macmillan. 1905.

Monroe, W. S. Educational measurement in 1920 and in 1945. *J. Educ. Res.,* 1945, **38**, 334–340.

Monroe, W. S., *et al. Ten Years of Educational Research,* 1918–1927. Urbana, Ill.: Univer. of Illinois Press, 1928.

Morris, N. L. *William James: the Message of a Modern Mind.* New York: Scribners, 1950.

Munn, N. L. The psychology of learning and its classroom application. *Peabody J. Educ.,* 1942, **19**, 257–265.

Ogden, R. M. *Psychology and Education.* London: Routledge, 1926.

Park, Dorothy G. Freudian influence on academic psychology. *Psychol. Rev.,* 1931, **38**, 73–85.

Rice, J. M. The futility of the spelling grind. *Forum,* 1897, **23**, 163–172, 409–419.

Sharp, Stella E. Individual psychology: a study in psychological method. *Amer. J. Psychol.,* 1899, **10**, 329391.

Thorndike, E. L. *Educational Psychology.* New York: Lemcke & Buchner, 1903.

Thorndike, E. L. *Elements of Psychology.* New York: A. G. Seiler, 1905.

Watson, R. I. The present status of educational psychology and educational psychologists. To be published.

Wissler, C. The correlation of mental and physical tests. *Psychol. Monogr.,* 1901, **3**, No. 6.

1.2 *Interpreting Research in Educational Psychology*

DONN BYRNE AND HENRY CLAY LINDGREN

Reprinted and abridged from Chapter 10, "Statistics and Measurement in Psychology," in *Psychology: An Introduction to a Behavioral Science,* 3rd edition, New York: John Wiley and Sons, 1971, by Henry Clay Lindgren and Donn Byrne, with permission of the authors and publisher. Donn Byrne is professor of psychology at Purdue University and is a leading figure in experimental social psychology and personality research.

This discussion of psychological statistics is included as an aid to students who may be perplexed by the statistical treatment of data in some of the articles in this book, as well as in other research studies conducted by educational psychologists. A major activity of educational psychologists is research, and it is impossible to have a clear understanding of their work without at least a "recognition vocabulary" of statistical terms. The terms discussed in this selection are the basic ones familiar to every psychologist.

It is by no means unusual to learn that a student responds to statistics as a source of concern and a possible barrier to further interest in the field of psychology. Part of the problem would seem to be that many of us have had unpleasant and not particularly enlightening experiences with mathematical courses in elementary school and beyond. By the time one reaches college, words such as "mathematics," "formula," or "equation" have acquired the power to evoke a negative reaction and even anxiety. College students faced with the prospect of taking a course in statistics (or even the task of reading a chapter such as the present one) often view the prospect as unpleasant, impossible, or both. With vast changes in the techniques of teaching and the introduction of the concepts of new math, this situation may not be a permanent one. It will be interesting to see whether college students in the near future are more relaxed and knowledgeable about the subject.

Before such changes become widespread, it might be helpful simply to learn that the subject matter of statistics need not be as nightmarish as students expect. In fact, we are already familiar with a great many statistical concepts. We read that a football team has *averaged* three touchdowns per game this season. We note that the *range* of yesterday's temperature was from a low of 58 to a high of 75 degrees. It can be seen that there is considerable *variation* in the size of adult human beings but relatively little *variation* in the size of adult St. Bernard dogs. Newspaper reports indicate that further evidence has been found to show that the incidence of lung cancer is related to the number of cigarettes smoked per day. We learn that women have a *longer* life span than men. We *sample* the cheese dip and decide whether to eat more. We are told that there is a 30 percent *chance* of rain today. Each italicized word suggests the use of an important statistical concept, which will be discussed in this chapter.

There are two major uses for statistics in psychology and education and both were represented in the examples above. First, statistics are used in *describing* events; some statistical terms serve to summarize a body of information and thus make it easier to communicate. Words such as "average," "range," and "relationship" are part of the statistics of description. Statistics are also used in drawing *inferences* about the meaning of data; when we wish to use our data to make predictions, there are statistical approaches by which we evaluate or estimate the accuracy of these predictions. Words having to do with "chance," "probability," or "sample" are included in the statistics of inference. In this chapter, both uses of statistical tools will be described along with the statistical evaluation of tests.

Statistics of Description

Describing a group of numbers: The frequency distribution

A first step in describing a series of numbers such as test scores, heights, or annual incomes is to convert them into a form that enables us to compre-

TABLE 1. Raw Scores on a Final Examination Made by Twenty-One Hypothetical College Students, Listed Alphabetically

Ashmore	75	Hamilton	65	Ortiz	51
Bailey	64	Iwabuchi	66	Parker	73
Cargill	46	Jackson	55	Reinhardt	65
DePue	71	Kazak	49	Sanders	84
Eilers	59	Leach	68	Thompson	40
Flores	57	McDonald	79	Wong	90
Goldstein	81	Neumann	65	Young	62

hend them as a whole. After a group of individuals (or nations, families, or whatever) has been measured with respect to some variable, the result is a series of numbers: the "raw scores." In order to illustrate the process of describing a set of measurements, we will utilize a hypothetical study. The same imaginary data will be used throughout the chapter in illustrating several statistical concepts. We will say that the individuals involved are 21 college students in a psychology class. The first data to be presented are their scores on a final examination.

The exam scores of this imaginary class are shown in Table 1, in the same way they might appear in an instructor's grade book. One glance at the 21 scores in Table 1 shows why it is that statistics of description are necessary. It is difficult, if not impossible, to make sense out of this array of names and numbers. Accordingly, the first step to be taken in describing the characteristics of a group of people on any variable is to make a frequency distribution, as shown in Table 2. Generally, it is convenient to divide the range of scores into 12 to 15 equal intervals, though a smaller number is used here for purposes of illustration. Then a tally is taken of the number of subjects whose

TABLE 2. Frequency Distribution of Psychology Final Examination Scores

Scores	Tallies	Frequency
90–95	/	1
84–89	/	1
78–83	//	2
72–77	//	2
66–71	///	3
60–65	~~////~~	5
54–59	///	3
48–53	//	2
42–47	/	1
36–41	/	1

scores fall within each interval. Although a frequency distribution is a somewhat imperfect description of the group's standing on the variable being measured, it is much easier to form a picture of the student's standing on the final by inspecting Table 2 than by trying to make sense out of Table 1. The actual way in which the distribution is presented may take a variety of forms. The same information contained in Table 2 is presented in two different graphic forms in Figures 1 and 2.

Describing the central tendency of the scores

One of the simplest, most familiar, and most useful ways of describing a group with respect to any variable is the

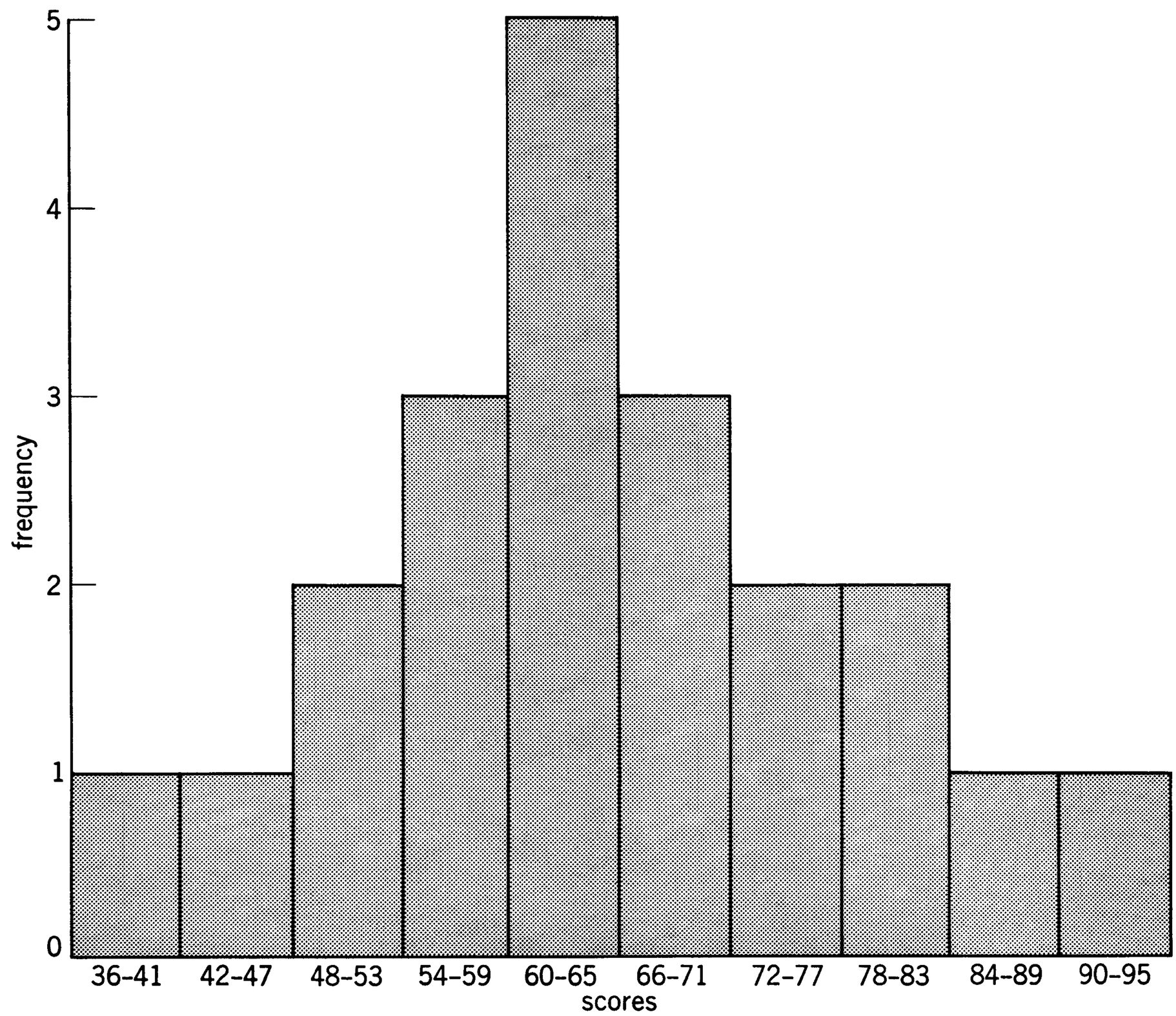


FIGURE 1. Ways of presenting data: histogram.

average. We speak of the average weight of American women, the average yearly rainfall in Los Angeles, and a student's grade point average. When we compute an average, it is almost always the *arithmetic mean* that is utilized. To obtain the arithmetic mean, we simply add all the scores and divide by the total number of scores. When the mean score is obtained, it gives us a convenient way to express approximately where a given group stands on a particular variable. In our example, the mean final examination score of the 21 students is 1,365 ÷ 21, or 65.

A thick book listing the height of every individual male in this country would be infinitely more difficult to comprehend than a sentence saying "The mean height of the American male is 5 feet, 9½ inches." Knowledge of the mean also makes comparisons possible. If we are told that the mean height of American women is 5 feet, 4 inches, we can see at a glance that men tend to be taller than women and by how much. We call the mean a measure of central tendency because it tells us something about where the center of the frequency distribution lies. It is thus a useful shorthand way of describing a group.

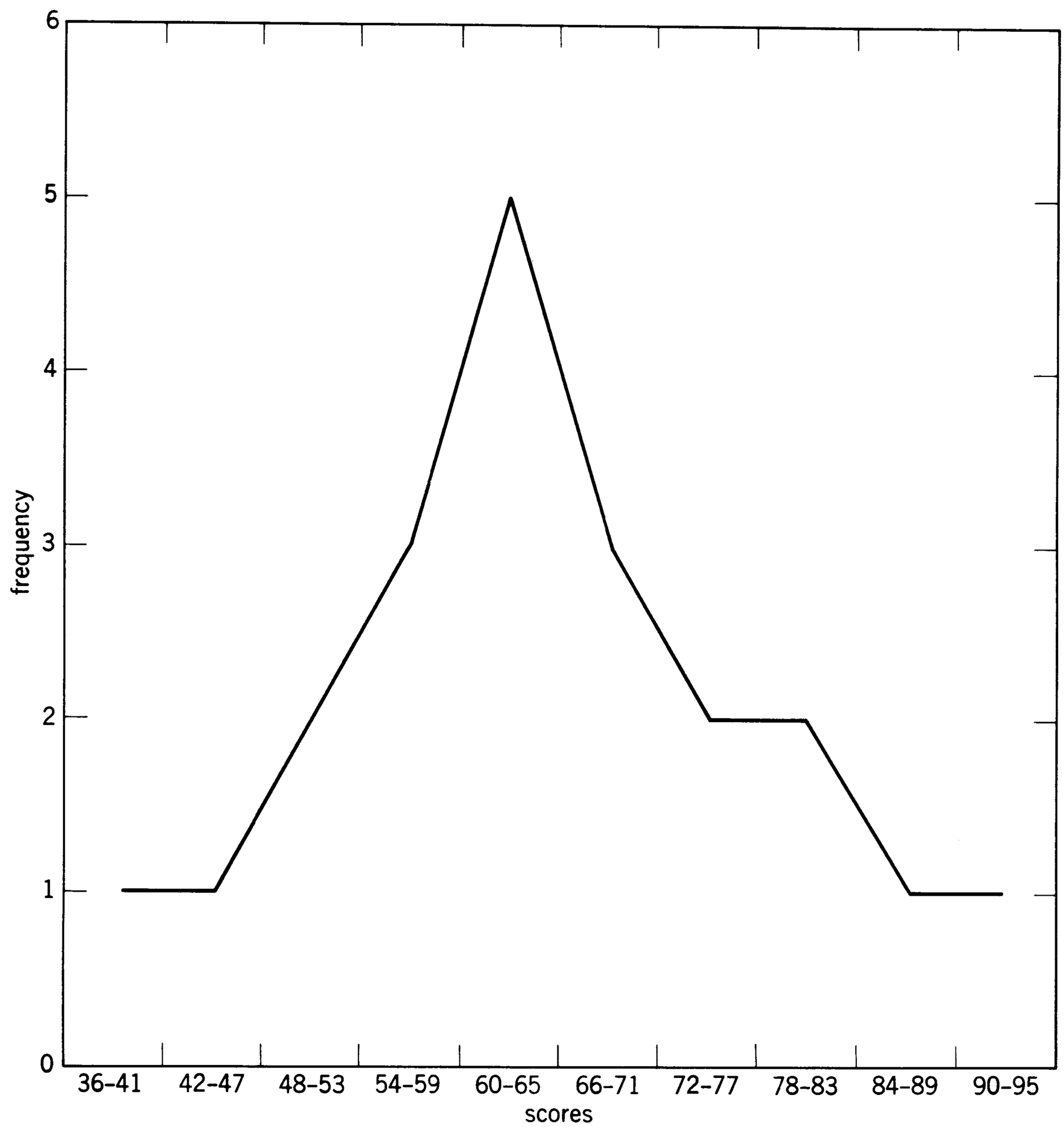


FIGURE 2. Ways of presenting data: frequency polygon.

Although the mean is the most commonly used measure of central tendency, two other measures are sometimes employed. The *median* is the mid-most score; half of the individuals in a distribution receive scores below the median, and half scores above the median. In most instances the mean and median are the same or almost the same score. That is, in the usual situation a group's average score and the midmost score tend to fall at about the same place. In our example, we found that the mean of the scores was 65; the median is also 65. In Table 1 it can be seen that three students received scores of 65, while 9 received lower scores and 9 received higher scores.

If the mean and median are often identical, why would the median ever

be used? It is because in certain circumstances the median gives a more accurate picture of the central tendency than does the mean. Such circumstances occur when a few extreme, unrepresentative cases are included in the distribution. As a hypothetical example, Table 3 presents the annual

TABLE 3. Annual Income of Residents of an Imaginary Gulf Coast Village (In Dollars)

1. 750	8. 1,800	15. 3,000
2. 900	9. 2,000	16. 3,100
3. 1,000	10. 2,000	17. 3,100
4. 1,500	11. 2,500	18. 3,500
5. 1,500	12. 3,000	19. 3,550
6. 1,750	13. 3,000	20. 3,600
7. 1,800	14. 3,000	21. 400,000

incomes of the 21 residents of a Gulf Coast village. The village houses a handful of poor fishermen, craftsmen, and tradesmen. In addition, resident 21 is a wealthy manufacturer who lives in semiretirement in a mansion overlooking the harbor. His annual income of $400,000 represents an extreme, unrepresentative case in this distribution. The town's mean annual income is a very comfortable $21,254.76 ($446,350 ÷ 21) but a more realistic figure is the town's *median* annual income of a meager $2,500 (the midmost score, which happens to be the income of resident 11). In similar situations, the use of the mean as an indication of central tendency is both inappropriate and misleading. It might be noted that this is one example of the way in which it is possible to misuse statistics to distort or misrepresent facts without actually lying. It is also an example of why it can be important for informed citizens to know something about the statistics of description.

A third, little-used measure of central tendency is the *mode,* or score obtained by the largest number of individuals. In most situations, the mode will not differ greatly from the mean and median. In Table 1 more students received a score of 65 than received any other score, so 65 is the mode. Sometimes the modal score is not representative of the distribution of scores, and there may be more than one mode. In the latter case the distribution is called bimodal, or trimodal, and so forth. An example of a bimodal distribution is shown in Figure 3. Our imaginary psychology class took a midterm examination, and the scores were distributed as shown in Figure 3. On this examination the midmost score or median is 57 and the mean is 59.4. Either of these figures is more representative of the central tendency than is the mode of 44.5 (the center of the interval from 42 to 47). All that the latter figure indicates is that there happens to be a bunching of scores at one point at the lower end of the distribution. The mode would almost never be used in preference to the mean or median but it is sometimes included as an additional descriptive device. When a distribution contains more than one mode, it is frequently the case that two distinct groups are present. For example, if our class consisted of some students who studied for the midterm and some who did not, it would not be surprising to obtain a distribution such as that in Figure 3.

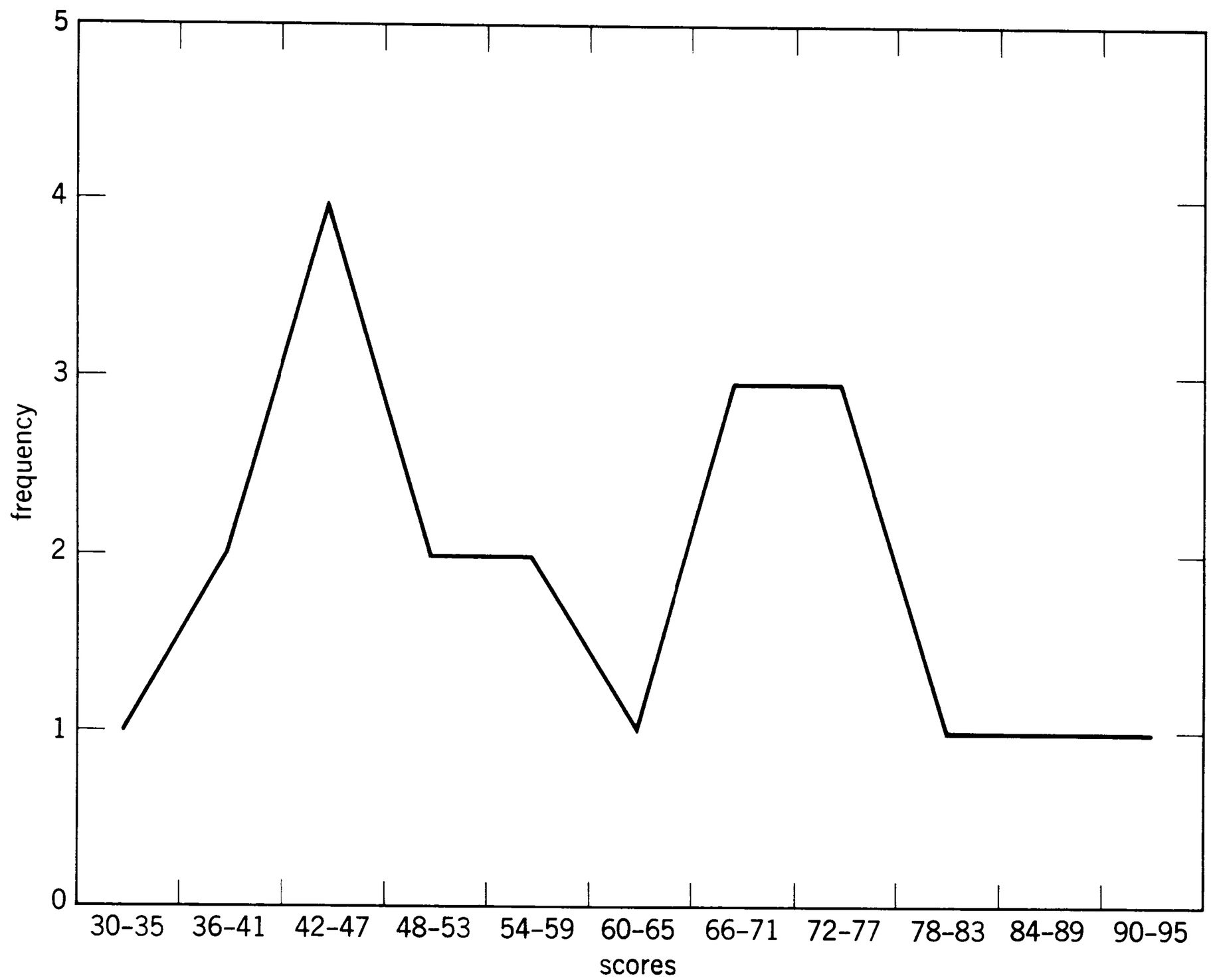


FIGURE 3. Bimodal frequency distribution: scores on a psychology midterm examination.

Describing the variability

Merely knowing the central tendency of a distribution of scores is not as informative as also having some knowledge about how scores are spread out around this point. The simplest measure of variability is the *range* from the lowest to the highest score. Thus, the final examination scores in Table 1 have a range of 50 (from 40 to 90), and the annual incomes in the Gulf Coast village in Table 3 range from $750 to $400,000 (a range of $399,250). Knowing the range between the lowest and highest scores gives us the area within which all scores fall.

Even the range leaves much to be desired by way of describing the group's variability because scores may fall in a variety of ways between the bottommost and topmost limits. They may bunch up around the mean or they may be spread evenly throughout the range, and one extreme score (as in the village) may extend the range far beyond the remainder of the scores. It would therefore be useful to know where most of the scores fall. The most useful statistic for this purpose is the

standard deviation. When the standard deviation is computed,[1] we then know much more about the variability of our scores. In a theoretically normal distribution, slightly over two-thirds—68.27 percent—of all the scores fall between the limits set by one standard deviation below the mean and one standard deviation above the mean. A total of 95.45 percent fall between two standard deviations above and below the mean. Almost all the cases—99.73 percent—fall between plus and minus three standard deviations.

The student may wonder as to the use of such information. First of all, the standard deviation is useful in understanding just how the scores of a particular group are distributed. When you learn that in the group of individuals on whom the Stanford-Binet Intelligence Scale was standardized (that is, the "standardization group") the mean IQ was 100 and the standard deviation was 16, you know a great deal about the distribution of IQ scores and the relative position of any given IQ score. A moment's calculation will show you that 68.27 percent of the standardization group (and presumably of the general population as well) had scores falling somewhere between 84 and 116 (100 $\pm$ 16), that 95.45 percent fell between 68 and 132 (100 $\pm$ 32), and that almost all IQs fell between 52 and 148 (100 $\pm$ 48). It can be seen that the mean and standard deviation are very important tools in describing and making sense of a distribution of scores. The second value of the standard deviation is that it makes the meaning of a particular score much more clear. An IQ of 132 is no longer just a number hanging in the air. With a statistical knowledge of the distribution you know that an individual attaining an IQ of 132 scores higher in intelligence than 97.7 percent of the population. Figure 4 shows a normal distribution of scores with points indicated where various multiples of the standard deviation would fall.

Returning to our final examination scores, the same sort of information may be utilized if we wished to administer that test to a variety of other students. (It should be kept in mind, however, that 21 students constitute a very much smaller sample and that they are not nearly as representative of the total population as was the group involved in the standardization of the Stanford-Binet.) We already know that the mean is 65; computation of the standard deviation yields a figure of 12.53. Can you determine between what scores on the test 68.27 percent of the students should fall? What scores should mark the limits for 95.45 percent of the students and for 99.73 percent? (Answers to these questions will be found at the end of this selection.) . . .

Statistics of Inference

Sampling

Behavioral scientists are continually seeking the answers to questions about behavior. Is massed practice more effi-

[1]The standard deviation is equal to the square root of the sums of the squared deviations (differences or D) of scores from the mean, divided by the total number of (N) of scores: S.D. $= \sqrt{(\text{Sum of } D^2)/N}$.

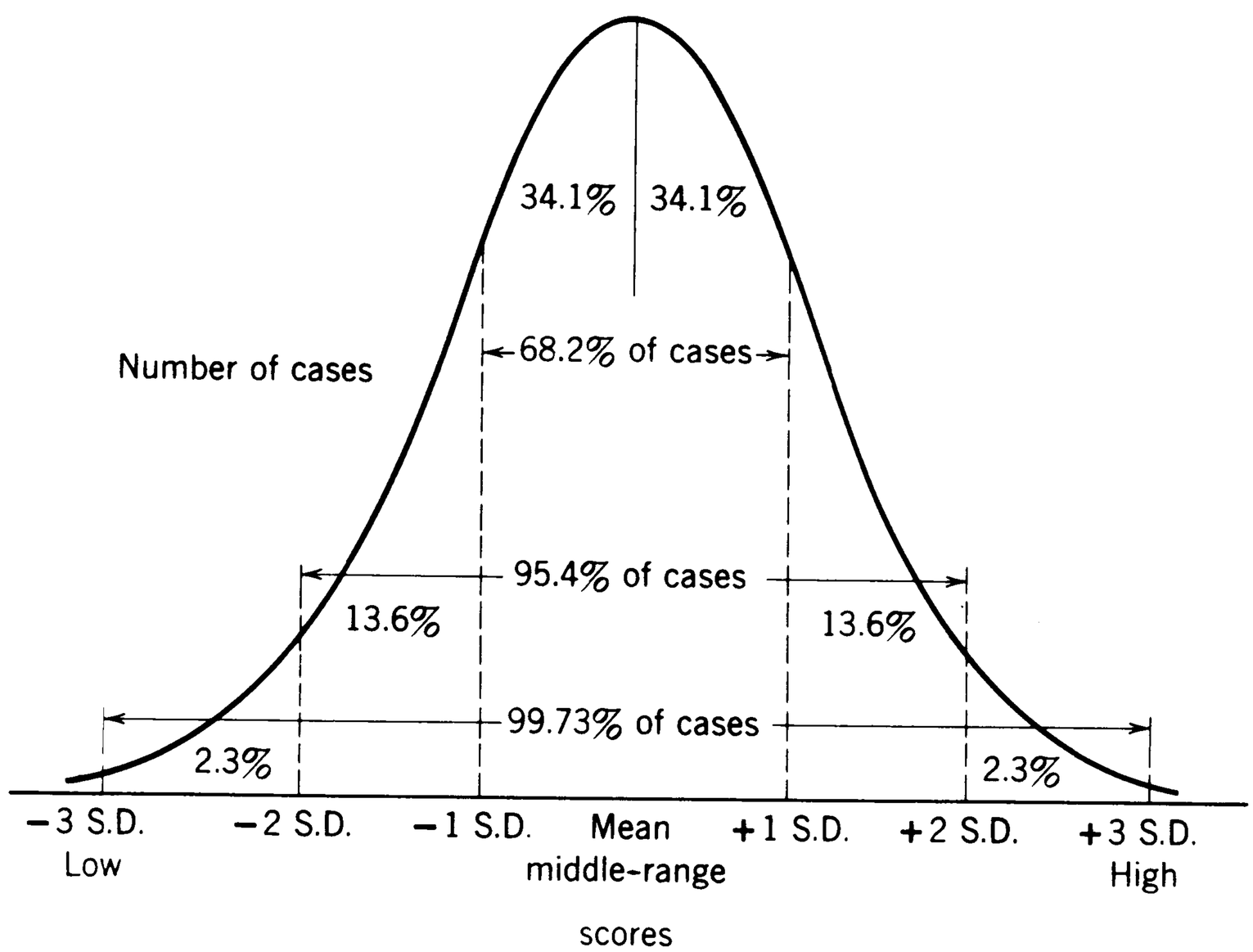


FIGURE 4. The standard deviation in a normal distribution of scores.

cient than distributed practice in learning? Do mothers of schizophrenic patients have different childrearing attitudes than mothers of nonschizophrenics? Is there a relationship between scores on an IQ test and grades in school? What is the relationship between stomach contractions and the subjective feeling of hunger? Does television violence have any effect on the aggressive behavior of viewers? Research is conducted with a group of subjects and then inferences are made about the larger group of individuals from whom the subjects were selected. If it were possible to study every human being in the world, we would not be concerned with the statistics of inference. We might find, for example, that distributed practice is twice as efficient as massed practice for the majority of the population and that would be that.

In reality, however, we never deal with the entire world's population and almost never deal with an entire group (such as *all* seniors in *all* colleges, or *all* Southern Baptists). Instead, *samples* are taken of various sizes from the groups in which we are interested. In *random sampling,* we try to choose subjects on a random basis and endeavor to select them in a manner that will eliminate bias. For example, in a study of the sexual behavior of college students, one would not want to have as subjects *only* those students who have sought help at the counseling center or *only* those students who volunteer to take part in a sex study or *only* students enrolled at a sexually segregated religious institution. Each of these potential samples constitutes a special

segment of the population and one that may well differ systematically from the general population in various ways. If, instead, we were to study every tenth student in the college directory at a variety of schools, we would have come much closer to obtaining a random sample of college students.

Another approach is to have the sample match known characteristics of the total population under study. This type of sampling procedure is called *stratified sampling* and is used by most professional pollsters, such as Gallup, Roper, and Harris. If you were interested in the television viewing of Maryland adults, for example, you might decide to build a stratified sample that would include the same proportions of individuals with certain characteristics, such as sex, age, marital status, race, income, area of residence, as will be found in the total population of the state. Within these limitations, the selection of subjects is on a random basis, of course.

As a further illustration, let us consider a study of television viewing. We will say that a potential sponsor manufactures an expensive product and wants to know whether individuals in a high income group are more likely to be attracted to a western show or to a science fiction show. Pilot films of each have been produced, and a choice must be made. To answer this question presents a problem because it is difficult to select subjects in a truly random fashion. Ideally, every adult in America whose income is over a certain level would be assembled around closed-circuit television sets, viewing both programs; then their preferences would be determined. This is, of course, impossible; hence the problem of sampling arises. It might be possible, although not feasible, to pick every one-hundredth or every five-hundredth eligible subject across the country. More probably, you would have to limit yourself to selecting some adults who live in your immediate area and attempt to sample them either randomly or on the basis of stratified characteristics. Suppose a sample of 500 individuals is obtained and it is found that the mean income of those who prefer the science fiction program is $20,000 and the mean income of those who prefer the western program is $6,000. Can you conclude that those who prefer science fiction are more affluent? What if the two means were $13,001 and $13,000; would your conclusions be the same? What if your sample had consisted of only four individuals and the two who preferred the science fiction show had a mean income of $20,000 and the two who preferred the western had a mean income of $6,000; would you be as sure of your findings as if you had studied the preferences of 100,000 subjects? If the range of the incomes of those who preferred science fiction was from $3,000 to $70,000 and of those who preferred the western was from $1,000 to $50,000, would there be so much overlap between the two groups as to make you less certain of the meaning of the differences between the mean incomes of the two groups? What if the ranges had been from $14,000 to $70,000 for one group and from $1,000 to $13,000 for the other; would the difference between the two means seem more impressive?

Basic to all these considerations is

the final question that must be answered: Would you find the same income differences related to program preferences if you repeated the study with another group of adults? In short, how sure can you be that your findings may be generalized beyond your sample? How sure can you be that your sample of subjects did not yield these findings simply by chance? The presence of any given subject in a particular sample is somewhat accidental. What if the 20 individuals with the highest incomes had failed to participate in your study for one reason or another? The means would be different from those reported above, perhaps even very close to one another. Thus, even if in the total population those with higher incomes really do prefer science fiction to westerns, this particular sample might have failed to reveal such a difference. In order to determine the likelihood that the findings in any one particular sample represent a real difference instead of simply a chance difference, we must turn to the statistics of inference and consider probability and significance level.

Probability

Although the mathematics of probability is a highly specialized field, almost everyone has some familiarity with the general idea. The chances are 50–50 (probability = .50) that when an unbiased coin is tossed in the air, it will come up heads. Figures show that 97.5 percent of the infants born in this country live to be a year old, so the chances are 97.5 out of 100 (probability = .975) that a particular infant will survive the first year of life. The weather bureau says that the chance of rain next Thursday is 40 percent (probability = .40). We are reasonably certain that if we hold a book in mid-air and let it go, it will drop to the floor (probability = .99 +).

If these probability figures are correct, it means that you should be able to make a number of fairly accurate guesses about the events involved. Your best bet would be that a coin tossed 4,000 times would come up heads 2,000 times, that 25 out of every 1,000 American babies will die before they reach the age of one, that one hundred such weather predictions will be followed by a rainy day 40 times and a dry day 60 times, and that the book will fall 300 out of every 300 times it is dropped. Psychologists make exactly these type of bets and guesses about whether experimental findings are applicable to the entire population. These bets or probability statements are based on certain characteristics of one's sample.

Significance level

The inferences that psychologists make about their experimental results with samples of the population are expressed in terms of levels of significance. If a psychologist reports that he performed an experiment and found that the mean income of those who preferred science fiction was higher than the mean income of those who preferred westerns at the 5 percent level of significance, he is indicating that the chances are only 5 out of 100 (probability = .05) that these same findings would not hold true in the total

population, and 95 out of 100 (probability = .95) that they *would* hold true. Another way to think of it is if in the total population those who prefer science fiction and those who prefer westerns really do not differ in income, 100 experiments with 100 different samples of subjects would still have found an apparent income difference of that magnitude only 5 times by chance alone. There is no way to know absolutely and positively, of course, whether in the general population the incomes of those who prefer the two types of programs *really* differ or whether this particular sample simply yielded a chance finding that is not true for the rest of the population. The best a scientist can ever do is make an estimate or bet. If the mean difference is found to reach the 5 percent level of significance, the odds are 95 to 5 (or 19 to 1) that the difference is a real one. If the difference reaches the 1 percent level of significance (probability = .01), the odds reach 99 to 1. At the 0.1 percent level (probability = .001), the odds are 999 to 1 that the finding is a real one. Obviously, the better the significance level, the more confident we can feel about the results.

It should be clear that "facts" in science are never a complete certainty. The repetition of experiments is of extreme importance, because duplicate findings in a new sample of subjects lend strong support to the original ones. If ten experimenters with ten different samples find that the mean income of those who prefer science fiction is higher than the mean income of those who prefer westerns, one could fairly confidently propose that this represents a real, nonchance difference in the viewing habits of different income groups. . . .

The Relationship Between Two Variables

Describing the relationship: The correlation coefficient

Among the most important and most frequently asked questions in psychology, as in other sciences, are those having to do with relationships. Is frequency of delinquent behavior related to average number of physical punishments received each month in childhood? Are the number of hours spent in practice related to the performance score later achieved in a certain task? Is there a relationship between the number of times television commercials are presented and the subsequent sales of the advertised products? Is there a relationship between racial prejudice and amount of education? The direction and magnitude of relationships are expressed by a single statistical measure, the *correlation coefficient.*

When we speak of the "direction" of the relationship, we are concerned with whether high scores on one variable go with high scores or with low scores on a second variable. A positive correlation indicates that high scores on variable X tend to go with high scores on variable Y, and that low scores on X tend to go with low scores on Y. For example, a positive correlation is usually found between IQ scores and school achievement: students with high IQs tend to get better grades than students with low IQs. A negative correla-

tion indicates that high scores on variable X tend to go with low scores on variable Y, and that low scores on X tend to go with high scores on Y. For example, a negative correlation is found between the number of hours of practice on a typewriter and the number of typing errors made on a test; after more practice, typing students tend to make fewer mistakes.

The "magnitude" of a relationship indicates the degree to which two variables are related. A perfect positive correlation is indicated by a coefficient of + 1.00; a perfect negative correlation is − 1.00, and 0.00 indicates a complete lack of relationship between the two variables. . . . In trying to understand the meaning of a correlation between two variables it may prove helpful to consider the relationship in terms of plots on a *scatter diagram.* Returning to our hypothetical psychology class, we shall examine the relationship between the scores they made on the final examination as given in Table 1 and those made on their final examination in an English course. In Table 4, the 21 students are listed with the final exam scores of each student in the

TABLE 4. Scores on Final Examination in Psychology, a Final Examination in English, a Midterm Examination in Psychology, a Performance Test in Physical Education, and Number of Hours Spent Watching Television on the Day Before a Final by 21 Hypothetical College Students

Student	Psych. Final	Eng. Final	Psych. Midterm	P. E. Performance	Hours of Television
Ashmore	75	84	70	46	0
Bailey	64	40	54	40	6
Cargill	46	49	36	65	9
DePue	71	79	75	51	10
Eilers	59	59	57	71	1
Flores	57	46	44	79	3
Goldstein	81	71	84	75	1
Hamilton	65	57	63	90	0
Iwabuchi	66	65	66	68	5
Jackson	55	65	43	84	12
Kazak	49	66	40	49	10
Leach	68	90	73	65	4
McDonald	79	55	67	73	2
Neumann	65	51	50	65	0
Ortiz	51	68	45	55	7
Parker	73	75	77	64	0
Reinhardt	65	62	47	81	6
Sanders	84	81	80	57	0
Thompson	40	64	30	59	7
Wong	90	73	93	62	0
Young	62	65	53	66	8

two courses plus additional data that will be discussed later. Figure 5 is a scatter diagram depicting the relationship between scores on the two finals. The Y, or vertical axis represents one variable (in this instance, the scores in English) and the X, or horizontal axis, represents the other (scores in psychology). Each dot represents the two scores of a single individual, indicating where he falls on each of the two variables. In our example, we have added the names of the individuals by the corresponding dots in order to clarify the process, but one would ordinarily use only the dots. When there is a high degree of relationship between the two variables, the dots fall close to a diagonal line. The less the pattern made by the dots resembles a diagonal line, the less the relationship and the lower the correlation. A diagonal pattern run-

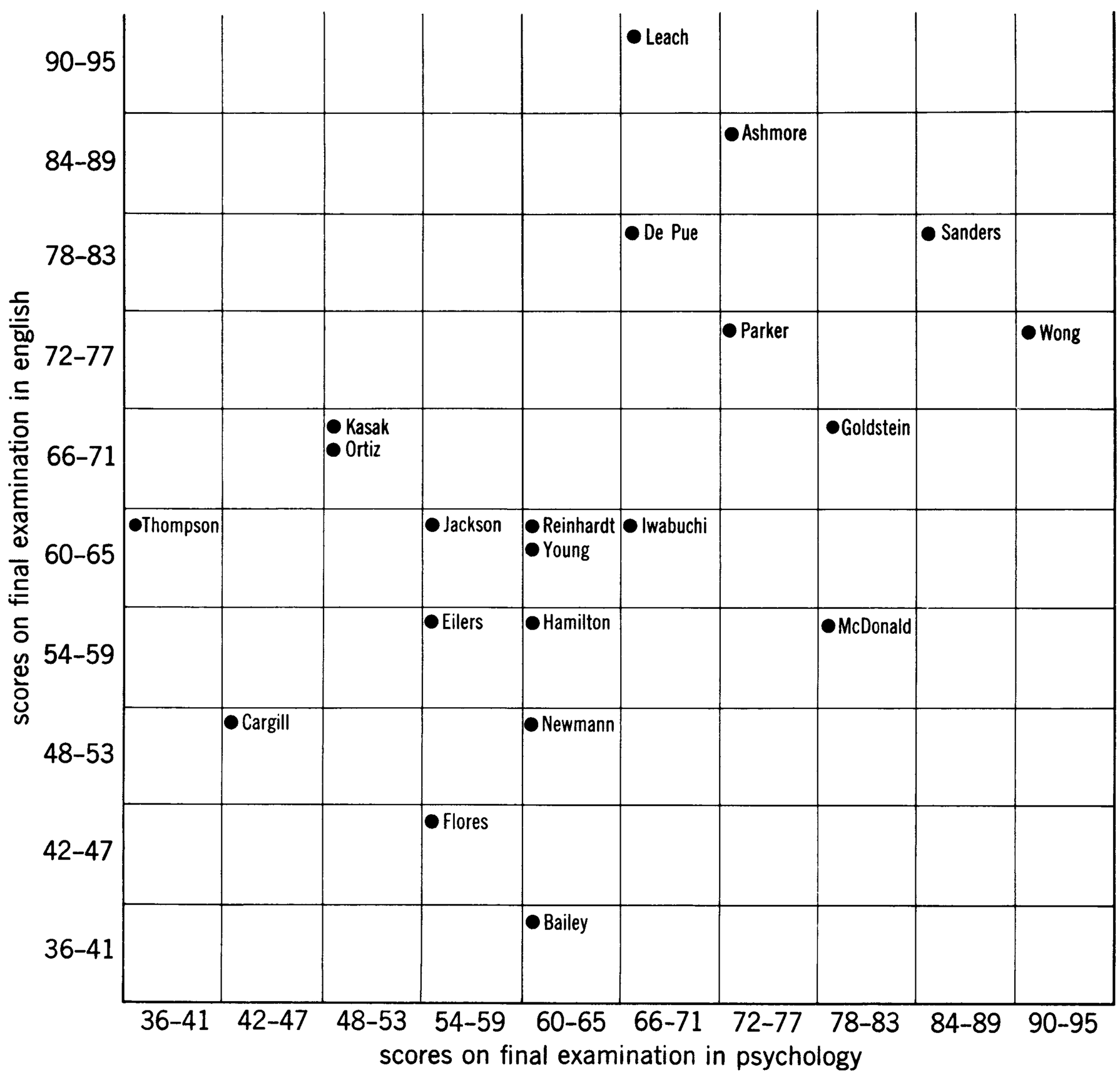


FIGURE 5. Scatter diagram showing relationship between scores on a final examination in psychology and a final examination in English. This relationship constitutes a moderate positive correlation (+ .42).

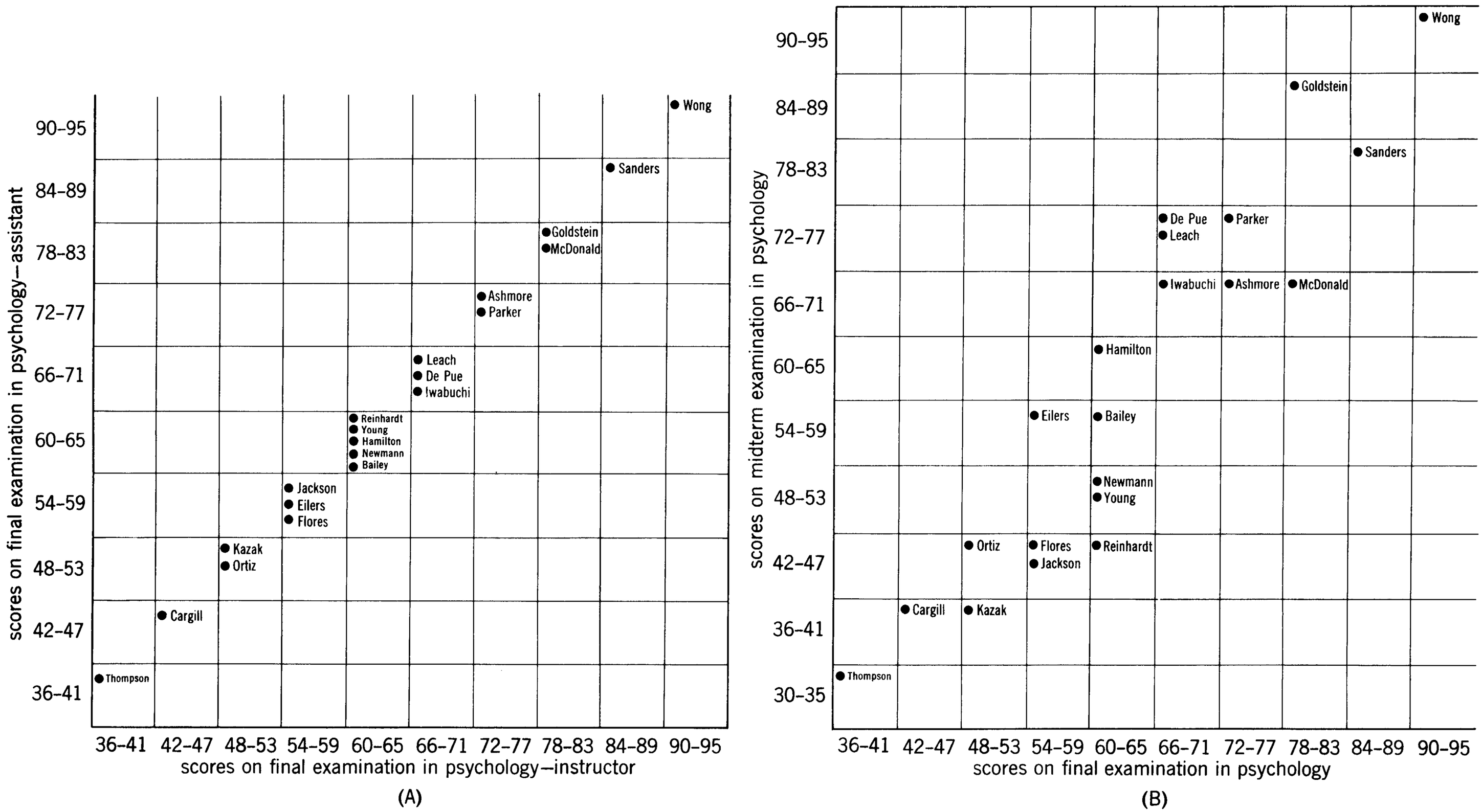
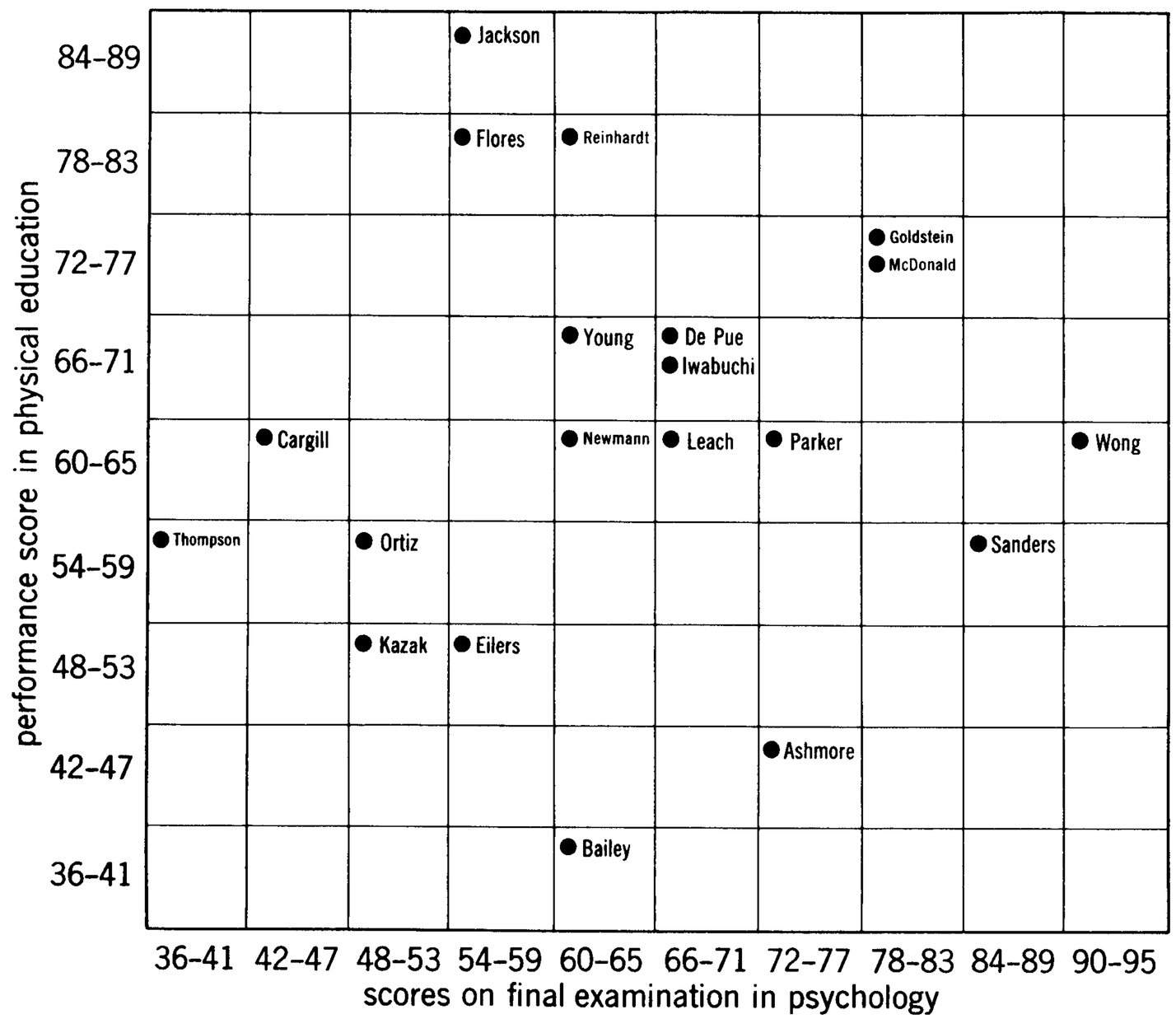
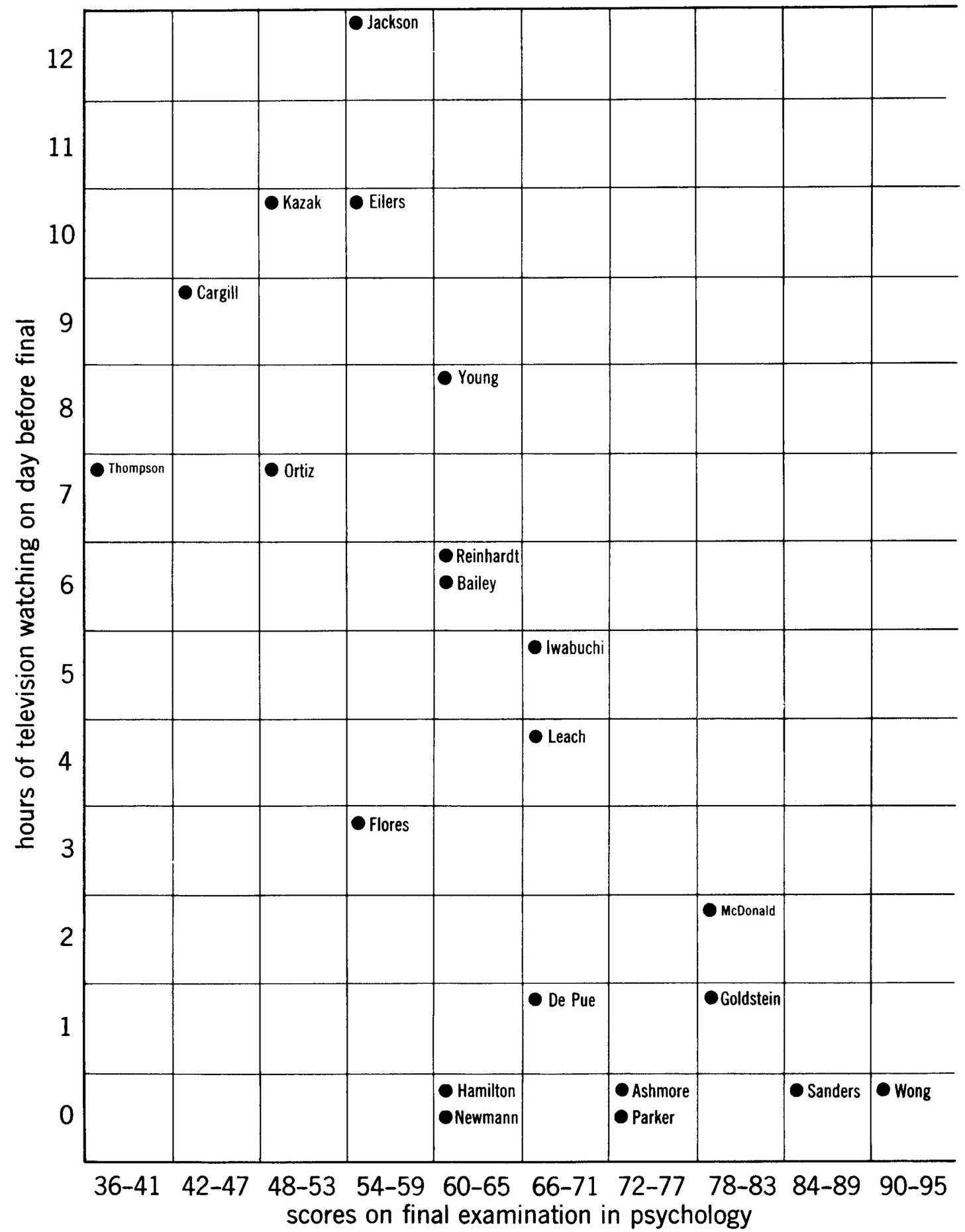


FIGURE 6. Scatter diagrams showing positive and negative relationships of various magnitudes for the group of 21 hypothetical college students. (A) Perfect positive correlation (+ 1.00). Scores on an "objective" final examination in psychology scored by the instructor and his teaching assistant. (B) High positive correlation (+ .93). Scores on a final examination in psychology and a midterm examination in psychology.



(C)



(D)

FIGURE 6 (continued). (C) No relationship (+ .07). Scores on a final examination in psychology and a performance test in physical education. (D) High negative correlation (− .75). Scores on a final examination in psychology and number of hours spent watching television on the day before the final.

ning from the lower left of the diagram to the upper right indicates a positive correlation; a diagonal pattern running from the upper left of the diagram to the lower right indicates a negative correlation. The relationship shown in Figure 5 is a moderate positive correlation as indicated by the correlation coefficient of .42. Several scatter diagrams demonstrating positive and negative relationships of various magnitudes between two variables are shown in Figure 6.

Knowing the correlation between two variables tells you whether it is possible to predict one from the other with better-than-chance accuracy; the higher the correlation (the closer to 1.00), the more accurate the prediction. For example, if the correlation between aptitude test scores and grades in college were + 1.00 (or even − 1.00), the prediction would be perfect; the test score of a college applicant would tell you exactly how well he would do in school. A correlation of zero between the tests and grades would mean that the tests would not help in predicting academic performance. Examine the scatter diagrams in Figure 6 to determine how a score on X would allow you to make a prediction about Y. Whenever we know the correlation between X and Y is greater than zero, we can use scores on X to predict scores on Y. Scores on a college entrance examination taken in the senior year of high school can be and actually are used by college officials to predict college grade-point averages. The amount of the variation in Y that can be predicted on the basis of knowledge about X may be determined by squaring the correlation coefficient. Thus, if X and Y correlate .50, 25 percent (.50 × .50 = 25 percent) of the variation in Y can be accounted for on the basis of a knowledge of X, while 75 percent of the variation in Y would still not be accounted for. If the correlation were 1.00, 100 percent of the variation in Y could be predicted; with a .00 correlation, 0 percent could be predicted.

A word of caution is in order at this point. There is a common tendency to assume that a high correlation between two variables implies that one variable *causes* the other. For example, we would very likely find a fair amount of correlation (perhaps + .25 or more) between annual income and the hour in the morning that individuals regularly arrived at work (that is, the later the hour, the higher the income), but we would be mistaken if we assumed that the amount of income was the *result* of the hour at which each person begins his work. In order to explain the relationship, we would need to find a third factor—in this case, the occupational level of work. This is the variable related to both income and the beginning hour of work; it accounts for the relationship between them.

Significance of the coefficient of correlation

Statistical inference is necessary in order to evaluate the correlation coefficient. If a relationship is found between two variables in a given sample of the population, what are the chances that this same relationship holds true for the total population? Again, the mathematical procedures will be omit-

ted, but it is sufficient to say that the significance level of correlation coefficients must be reported by experimenters. How large must a correlation be in order to reach significance at the 5 percent level, or at the 1 percent level? The answer depends on the number of subjects in the sample in which the correlation was computed. Table 5 indicates the magnitude that a correlation coefficient must reach in order to be significant at the 5 and 1 percent levels. It may be seen in the table that the larger the sample, the smaller the correlation coefficient may be and still reach statistical significance. For example, with 3 subjects, the correlation must be perfect before the coefficient reaches the 1 percent level of significance. Yet, with 1,000 subjects the 1 percent level of significance is reached with a correlation coefficient of only .08. Of course, a correlation as small as .08 is not very useful for prediction; the fact that it is statistically significant simply tells us that it is probably a real relationship, although a slight one.[1]

TABLE 5. Numbers of Subjects Needed for Coefficients of Correlation to Attain Significance at the 5% and 1% Levels

Number of Subjects	Size of Correlation Coefficient Necessary to Reach 5% Level of Significance	1% Level of Significance
3	.997	.999
5	.88	.96
10	.63	.76
20	.44	.56
30	.36	.46
40	.31	.40
50	.28	.36
100	.20	.26
500	.09	.12
1,000	.06	.08

[1]Answers to questions: 52.47 and 77.53; 39.94 and 90.06; 22.41 and 102.59.

Van Bucher

PART 2

Growth and Development of Learners

2.1 *The Beginnings of the Self: The Problem of the Nurturing Environment*

IRA J. GORDON

Reprinted from the *Phi Delta Kappan,* 1969, **50**, 375–378, with permission of the author and Phi Delta Kappa, Inc. Ira J. Gordon is professor of education and director of the Institute for Development of Human Resources, College of Education, University of Florida at Gainesville.

Understanding the learner's behavior, and particularly his problems and difficulties, is greatly facilitated if we can see how he regards himself: his self-concept. Gordon traces the course of self-concept theory as it was conceived first by Descartes and later by Freud and more recently by psychologists in the mid-twentieth century. Today we are much concerned about our failure to educate children from poverty homes. An understanding of self-concept development during early childhood shows that the problem is, to a large extent, a motivational one and suggests, in broad terms, what we must do about it.

The concept of the self is an old one in religion and philosophy and has been discussed endlessly as a part of man's search for identity, as he sought to answer the question, "Who am I?" For Descartes the answer was, *"Cogito ergo sum"*—I think, therefore I am. This statement marked a sharp break with medieval thought, and contributed to the age of reason. For Descartes, cognition or reason was superior to emotion. Knowing was the self's primary function. The self was active, aware, free; the senses and emotions were passive, or confused influences upon the mind. From the early seventeenth until the late nineteenth century, this view reigned.

Freud broke with this tradition by centering upon the emotions, by denying free will, and by focusing upon the influence of the child's experiences in the earliest years. Since Freud, the Descartian answer is insufficient. We now seek to define ourselves in ways which include our feelings as well as our thoughts, and look for the origins of our personality in the first dim moments of life long before cognition seemed possible. Because of Freud, our notion of self-definition has required that its origins be in early childhood and that it be developed from the experiences we have had in that most intimate of circles—our family. For modern man, this is a truism; but it also leaves unanswered a myriad of questions concerning how we got that way.

The first step in self-awareness is both affective and cognitive: the discovery of one's own body as distinct and pleasurable. When the infant puts thumb in mouth, he experiences sensation in both his thumb and his mouth and learns that the thumb is part of him. When the numerous other objects that the infant places in his mouth do not yield the double sensation, he separates self from other. This process, labeled "self-sentience" by Sullivan,[1] provides the infant with his first anchorage point, his first awareness of separateness. To paraphrase Descartes, if the infant could speak, he might say, "I experience me, therefore I am."

The second marking point is the awareness of "other." The separation of "I" or "me" from "not me" requires the introduction of people and objects from outside the child. The child needs enough of them, with enough frequency and consistency, that they can be differentiated. The infant at three months engages in social smiles,[2] but much has gone on before this time to enable the child to reach this major social event. It is not purely the "maturation" of an inadequate organism toward social behavior. William James, at the turn of the century, defined the world of the infant as a blooming, buzzing confusion, but current research in learning indicates that infants are able to make much more elaborate differentiations of their physical environment in terms of sight and sound and sense than James would have thought possible.

For example, Lipsitt's research at Brown[3] indicates the ways in which both operant and classical conditioning can occur in infancy. Although its approach is not psychoanalytic, current research in infant learning substantiates the psychoanalyst's view of the infant's ability to learn and thus supports the notion of the importance of this early

period. But what is it the child learns, in addition to such behaviors as feeding or cooing responses, smiles, and cries? The period of infancy has been seen as the time the child learns basic trust.[4] The nature of the inputs—that is, the way he is handled and fondled, dealt with and responded to, and how his body reacts to these events—teaches the child whether or not the world is a safe or terrifying place, and whether he can trust it or not.

Since the separation of self and world is incomplete, the self-concept, the "I," is part of the world. It is both cognitive and affective, active and passive. "I" is not only in the brain but also in the viscera. It is both Cartesian and Freudian.

Specifically, what are some of the inputs in the very early years which influence the initial picture of the self? Robert Sears and his colleagues,[5,6] in a series of studies which applied learning theory rigor to psychoanalytic concepts, indicated that parental attitudes and behavior (disciplinary techniques, permissiveness, severity, temperamental qualities, and aspirations) exhibited in the areas of hunger, elimination, dependency, sex, and aggression were important factors in development and in sex-role identification, a major dimension of the self-concept. But these external inputs emphasize the affective side of life. They do not adequately consider either the cognitive dimension or the role of the child himself.

Current thought emphasizes the competence of the infant and brings together both the cognitive and affective elements of the child into one system. It emphasizes the importance of not only the characteristic child-rearing patterns described by Sears, and the family drama so dear to the psychoanalyst, but also the role of the infant himself as an active, striving, curious, learning organism who makes his impact on his family. This is no *tabula rasa* child. And the child's view of himself is not simply a mirror image of the external events which surround him early in life. From the very beginning it includes his own organism as it senses, feels, learns, and assigns meaning to these external stimuli. The child learns who he is from what happens to him, from the language that surrounds him, from the people who are dear to him, from the opportunities to deal with the objects and events in his immediate world, and from his own responses to the welter of stimuli. His self-esteem represents his unique organization of his own biological makeup, the evaluations made of him by significant adults, and his own learning from trial and manipulation and feedback from his world. Cognitive development is inseparable from personality development.

The child obviously cannot define "self" as distinct from "other" before he has a permanent frame of reference. One measure of this frame is Piaget's "object permanence," manifested by the individual's recognition that an object continues to exist even though it is no longer visible to him. He arrives at this point somewhere in the second year of life. It is a growth marker because now he can relate affectively to other individuals in some consistent fashion, and cognitively he has achieved a level where he can actively engage in searching his environment. Gaining this ability

is a giant step forward and gives the child a sense of competence in relating to his world.

We can make an intuitive leap from object permanence to Erikson's basic trust. Both mean that the child has now organized at least a portion of his world so that it is orderly and predictable—and therefore manageable. With this he can structure a positive self-concept. Without a sense of object permanence, he is powerless. Psychological inputs are important here, because only on the basis of broad experience can the child discover that both people and things have external reality. With the establishment of "other," the child's own behavior can now include role-taking and role-playing.[7] This process enables him to shift from Piaget's "egocentric" stage toward "decentration." That is, he develops from seeing others as just like him toward a recognition of the fact that what one sees and believes depends upon where he stands and what he already knows. Parents not only influence opportunities for such role-playing, but also provide the basic models for imitation. Through the ways in which they teach or deny opportunities for dramatic play, they influence both the cognitive and affective dimentions of the self-concept. Smilansky[8] has described the way parents affect this phase of learning.

Piaget's theoretical exposition of cognitive development returns us to the epistemological position of Descartes, but with added knowledge from Freud and the behavioral scientists. Decarie[9], for example, was able to investigate both Piagetian and psychoanalytic views about the process, timing, and meaning of arrival at object permanence. Generally, she found empirical support for both, and concluded that parents are the most effective agents in presenting both cognitive and affective experience to the young child. Piaget wrote in 1954:

> *The other person is of course an emotional object to the highest degree but at the same time is the most interesting cognitive object, the most alive, the most unexpected. . . . The other person is an object which implies a multitude of exchanges in which cognitive as well as affective factors play a role, and if this object is of paramount importance in one of these respects, it is, I think, equally important in the other.*[10]

How important are these early years? Gardner Murphy has indicated that the self-picture is fairly well integrated by the third year of life. Once it has developed, it becomes the evaluator, selector, judger, and organizer of future experience, and the child's behavior may be seen as organized to enhance and maintain his view. Such a picture sounds harsh and deterministic if we did not understand that possibilities for change are always present. Life is not over at age three, but the general view toward the world and toward one's self is already present.

The longitudinal data which support the importance of early childhood are fairly consistent. Bloom indicated on the basis of reviews of longitudinal research[11] that half of what accounts for the variance in adults in aggressiveness in males and dependence in females seems to be present by age four. Not

only Bloom's summary but also the classical longitudinal studies conducted in California[12] and the longitudinal studies of the Fels Institute at Yellow Springs, Ohio,[13] demonstrated the effects of parental behavior in the child's first six years on his behavior and attitudes in subsequent years.

One of Bayley's findings is that the mother's affectional behavior toward her son in the first three years of his life was related to his friendship, cooperation, and attentiveness when he became a school child and an adolescent. These behaviors may be inferred to be reflections of feelings of security, a fundamental dimension of self-concept.

One of Freud's contributions is the concept of identification, and, more specifically, sex-role identification. We noted earlier that Sears adopted this concept and applied general behavior methodology to its investigation. It is central also to Kagan and Moss. For them, the notion of sex-role identification is a core concept in influencing stability of behavior from childhood through adulthood. Events early in life lead not only to the child's sex-role identification but also determine his general social expectancy for all behavior. Boys are expected to behave more aggressively, more competently, and in more task-oriented fashions; girls are to be more nurturant, more person-oriented. Parent behavior in the first six years of life influences the child's identity and the standards he will set for typical sex-related behavior. Kagan and Moss conclude that the individual's own desire to make his behavior agree with the culture's definition of sex is a major factor determining the stability of his behavior over time.

Longitudinal studies indicate how very important it is to analyze data about children by sex as well as by age. They indicate the differential effects of parental behavior on boys and girls. This should not surprise us, but it often gets overlooked. In both the cognitive and affective aspects of the self, boys and girls view themselves differently, tend to use different learning styles, tend to evaluate different aspects of self and world as important. The origins lie both in biology and in differential treatment.

Unfortunately, most of the children studied in longitudinal research have been middle-class, from somewhat stable families, where conditions might generally foster the mix of intellectual and emotional inputs that lead to positive views of the self. They fit Lois Murphy's observation that "Each experience of mastery and triumph sets the stage for better efforts in the next experience. Confidence, hope, and a sense of self-worth are increased along with the increase in cognitive and motor skills, which can contribute to better use of the resources."[14] Their world provides them with both intellectual challenge and emotional support. Both the cognitive and affective "matches"—the connection between the child's motives and cognitive level on the one hand and the experiences being offered to him on the other[15,16]—are in phase. His positive self-image receives verification from his competence in dealing with the world.

Unfortunately, not all children have the sense of triumph described by Lois Murphy, nor do their selves match the

world's demands. Yarrow's studies[17] of maternal deprivation indicate the difficulties encountered by children who lack a mother figure to provide them with some stable anchorage points. In the social domain, Clark,[18] Deutsch,[19] Smilansky,[20] Marans,[21] and Wortis,[22] among others, point out the devastating effects of social deprivation on building positive self-esteem. Although the child's view of himself does not mirror and is not an exact replica of his world's picture of him, for many youngsters it comes quite close. If the larger society conceives of the child as not worthwhile and demonstrates consistently to him that it so judges him, it is difficult for the child to value himself. Children in the ghetto, children classified as slow learners, children who for a variety of reasons are told even in these early years that they are not quite good enough or smart enough or handsome enough tend to devalue themselves and thus to set the stage for continuously poorer levels of performance than might otherwise be their lot. These images are already set before entry into school.[23] Children growing up in psychologically disorganized homes suffer similar fates, as Pavenstedt[24] has indicated about South Boston children.

"As the twig is bent . . ." has long been part of Western folklore. Scientific data now support this view. The origins of the self lie in the early years. How the child will see himself is influenced by the way he is treated, the opportunities provided for him, how he is evaluated as he copes with these opportunities, and how he perceives these evaluations.

If these early years are crucial in determining school performance through the mechanism of the self-concept, then society cannot shrug off its responsibility. For very young children, negative self-views may be as damaging as physical illness or actual physical handicap. We are rapidly making provision for medical help. We need to create nurturing environments early in life so that children's concepts of themselves may possibly emerge as positive. Whether the school systems as now constructed are the appropriate agencies to reach down to the younger years is open to debate. The example of Head Start programs and the present Parent and Child Center movement indicate that new social agencies consisting of and requiring the participation of those for whom the service is intended may provide effective vehicles for change. What is needed is education so designed that parents can provide children not only with an *affective* climate which tells them they are loved and worthy but also with a *cognitive* climate that allows the child to be competent as well as feel loved. Adequate self-esteem requires this combination.

A characteristic of the American society is its own self-concept that it is capable of solving the problems which afflict it, once the problems are pointed out. The issue is clear. What is required now are social engineering skills. Intervention is essential. We have some ideas of what it should be and who should render it. Now we need to develop the types of programs which provide for all children the psychological inputs which lead to positive self-esteem.

References

1. H. S. Sullivan. *The Interpersonal Theory of Psychiatry.* New York: W. W. Norton, 1953.
2. René Spitz. *The First Year of Life.* New York: International Universities Press, 1965.
3. L. Lipsitt. "Learning in the Human Infant," in H. W. Stevenson, E. H. Hess, and H. L. Rheingold (eds.). *Early Behavior: Comparative and Developmental Approaches.* New York: Wiley, 1967, pp. 225–48.
4. E. Erikson. *Childhood and Society.* New York: Norton, 1951.
5. Sears, E. Maccoby, and H. Levin. *Patterns of Child Rearing.* Evanston, Ill.: Row Peterson, 1957.
6. R. Sears *et al., Identification and Child Rearing.* Stanford, Calif.: Stanford University Press, 1965.
7. G. H. Mead. *Mind, Self and Society.* Chicago: University of Chicago, 1940.
8. S. Smilansky. *The Effects of Sociodramatic Play on Disadvantaged Pre-school Children.* New York: Wiley, 1968.
9. T. Décarie. *Intelligence and Affectivity in Early Childhood.* New York: International Universities Press, 1965. (Translated by Elisabeth and Lewis Brandt).
10. J. Piaget. *Les Relations Entre l'Affectivé et l'Intelligence Dans la Développement Mental de l'Entant.* Paris: Centre de Documentation Universitaire, 1954.
11. B. Bloom. *Stability and Change In Human Characteristics.* New York: Wiley, 1964.
12. N. Bayley. "Consistency of Maternal and Child Behaviors in the Berkeley Growth Study," *Vita Humana,* 1964, pp. 73–95.
13. J. Kagan and H. Moss. *Birth to Maturity.* New York: Wiley, 1962.
14. L. Murphy and associates. *The Widening World of Childhood.* New York: Basic Books, 1962.
15. J. McV. Hunt. *Intelligence and Experience.* New York: The Ronald Press, 1961.
16. Ira J. Gordon. *Studying the Child in School.* New York: Wiley, 1966.
17. L. Yarrow. "Separation from Parents During Early Childhood," in Martin L. Hoffman and Lois W. Hoffman (eds.). *Review of Child Development Research,* Vol. 1. New York: Russell Sage, 1964, pp. 89–136.
18. K. Clark. *Dark Ghetto.* New York: Harper & Row, 1965.
19. M. Deutsch and associates. *The Disadvantaged Child.* New York: Basic Books, 1967.
20. Smilansky, *op. cit.*
21. A. Marans, D. Meers, and D. Huntington. "The Children's Hospital in Washington, D.C.," in Laura L. Dittmann (ed.). *Early Child Care, the New Perspective.* New York: Atherton, 1968, pp. 287–301.
22. H. Wortis, *et al.,* "Child-rearing Practices in a Low Socio-economic Group," *Pediatrics,* 1963, pp. 298–307.
23. B. Long and E. Henderson. "Social Schemata of School Beginners: Some Demographic Correlates," in *Proceedings,* 75th Annual Convention, American Psychological Association, 1967, pp. 329–30.
24. E. Pavenstedt (ed.). *The Drifters.* Boston: Little, Brown, 1967.

2.2 *Parental Expectations for Independent Behaviors and Achievement of Elementary School Boys*

ANITA WHITING

A previously unpublished paper delivered at the April, 1970, convention of the Western Psychological Association in Los Angeles, revised by its author specially for this book and printed with her permission. Anita Whiting has had considerable experience as a classroom teacher and a school psychologist. At present she is School Psychologist, Covina Board of Advisors, Perceptual Learning Center and Consultant Psychologist, Crippled Children's Society, San Gabriel and Los Angeles.

Although each child's self-concept is unique, it is strongly influenced by parental attitudes. In this report, Whiting shows how the school success of two groups of boys was significantly related to the expectations that their parents had for them. In general, parents of normally achieving boys, in contrast to those of underachievers, expected them to be independent, autonomous, and to be able to handle everyday problems without much help. They also tended to express democratic rather than authoritarian values in their relations with their children.

Background

Independence is a concept which has received far less than its fair share of theoretical and practical attention during the last 20–25 years. Psychologists have been concerned, perhaps overly concerned, with dependency and have seldom dignified "independence" by systematic discussion—much less by definition. As most frequently used, the term is intended to mean, simply, less dependence.

A change in orientation is under way, however. Thanks to Adlerian psychology and to individuals such as Erik Erikson (1950) and Robert W. White (1959) who stress the importance to a child of developing a sense of accomplishment and competence, behavioral scientists are taking a fresh look at independence and its related areas of competence and mastery.

At Claremont Graduate School, interest has focused on comparison of *expectancies* for independent behaviors as such are held by categorically distinct groups. It has been found, for example, that parents of patients at Casa Colina Rehabilitation Hospital consistently believed their children should be able to master behaviors at an earlier age than did the hospital's rehabilitation team (Keith, 1966). This difference was particularly apparent with respect to intellectual functioning; the lower the child's ability, the greater the parents' tendency to overestimate. My colleague, Dr. Audrey Gray (1968) who designed the Age–Independence Scale, or AIS, revised for the research to be described, found significant differences between the expectations of lower and middle class mothers. Middle class women anticipated significantly *earlier* mastery by their children of common tasks and activities across all of her scale categories.

It is the purpose of this paper to acquaint you with a study which had as its primary purpose investigation of the relationship between parents' expectations for independent behaviors and their sons' actual school performance. A secondary interest of the research study was the gathering of information on those child-rearing practices which seem to be especially relevant to a child's success or failure in school, e.g., the parents' attitudes toward homework and active involvement in their son's school assignments.

Psychologists working in the public schools have become increasingly concerned about these intellectually able children who consistently fail to perform at a level commensurate with their ability; who seldom undertake—much less complete—assignments without additional prodding from the teacher; who always manage to lose homework, promise to do better—but never do. Academic difficulty, for whatever cause, is presently the reason for the clinical referral of at least three-quarters of the children between seven and fourteen years of age (Kessler, 1966). Underachievement is also the single most common characteristic of emotionally disturbed children (Hobbs, 1966).

Research strongly suggests that underachievement in otherwise bright and able youngsters, children free from handicapping physical, e.g., perceptual–neurological, conditions, is a function of the total family; that the pattern of

behavior which results in academic failure probably begins *before* the child enters school—in his home and social setting. For example,

1. Social class is a major factor.
2. Underachievement is frequently associated with overdependence upon the mother (Levy, 1943).
3. Failure in boys of average intelligence begins very early. Studies show significance at the .05 level as far back as the first grade (Shaw & McCuen, 1960).

Thus, in an effort to examine one aspect of family interactions thought to have bearing on academic success or failure, the primary intent of this study was to determine whether parents of boys successful in the elementary school program viewed independence training in a different manner than did parents of boys who were unsuccessful. Did they hold different expectations as to the ages at which their children should have mastered certain tasks or developed particular skills? Did they differ with respect to the particular activities they thought most important for their sons to master independently?

Participants

Participants were 52 middle-class couples (104 individuals) from the same elementary school, chosen because one-half of them, i.e., 26 mothers and 26 fathers, had sons recognized as chronic nonperformers; the others were parents of boys successfully completing the academic portion of the elementary program. The study was limited to boys because underachievement is essentially a male problem (Bentzen, 1963); also as a result of the difficulties which would have been encountered in matching and the error that might have been introduced by comparing parental expectancies for both boys and girls. All of the youngsters were average or above in intelligence. The average I.Q. for the successful group was 114 and for the unsuccessful group it was 113. The boys selected as doing well had a grade point average that was consistently A or B for academic areas and study skills. Performance on group tests of achievement was at or above grade level. They were described by teachers as showing initiative, completing work, and functioning independently. Their comparison group, on the other hand, had a grade point average that was D or F and their measured achievement fell substantially (a minimum of six months) below grade level. These boys were seen as demanding attention, being unconcerned with failure, seeking help, and not completing assignments. Failure to try was a major characteristic of behavior.

The parents of these youngsters were seen together for approximately an hour and a half either in their home or in the psychologist's office. Interestingly, none of the parents whose sons were nonperformers, hereafter for brevity's sake referred to as unsuccessful parents, wanted the examiner to come to their home, while the group to be called successful parents usually elected their homes over the office. During the hour, mothers and fathers each, *individually,* completed the Age–Independence ques-

tionnaire. Then they participated, *together,* in a standardized interview designed for the project.

Of particular interest were the responses of participating fathers. In the area of child development, there has been relatively little theory and much less research on the father's role in child rearing. It was felt that fathers contribute importantly to their children's development—particularly in those areas which most directly reflect the child's competence, mastery, and independence.

Research and Methodology

The major research tool, the format of which is shown as Figure 1, was the Age–Independence Scale (AIS). Developed by Gray from the earlier work of Winterbottom and later revised for the present study, the Age–Independence Scale or AIS lists 110 common tasks and activities divided into six conceptually distinct categories consisting of items which pertain to Self-Care, Cognitive Facility, Physical Skills, Social Responsibility, Autonomy, and Wide Experience. The respondents replied to individual items by writing down the age at which each believed *his* son should have mastered the activity. In addition, each parent was asked to check those 25, and only 25, scale items which he considered to be the "most important" from among those listed. Major hypotheses were investigated by means of two-way analyses of variance with chi-square and Anderson's U-Statistic used to evaluate parental values and child-rearing practices.[1]

Findings

Major findings are summarized below.

1. The analysis of the parents' responses showed that there was a highly significant difference (at $p < .001$ on four of the six scales and at $p < .05$ on one) between parental groups with respect to ages at which they expected independent behaviors. In all cases, parents of successful boys expected mastery *earlier* than did parents of unsuccessful boys. The difference between groups was particularly apparent for

[1]Details of the statistical analysis may be obtained by writing the author at Claremont Graduate School, Claremont, Calif. 91711.

FIGURE 1

(Check 5 Only)	Age My Son Should	A More Important Item to Me
1. Bathe or shower self completely	(....)	(....)
2. Read a simple story	(....)	(....)
3. Make change with money	(....)	(....)
4. Swim length of public pool	(....)	(....)
5. Stay overnight with a friend	(....)	(....)
6. Ride bicycle in street with light traffic	(....)	(....)
7. Have a regular allowance	(....)	(....)

the scale of Autonomy on which the successful parents expected competence about one year earlier than did the unsuccessful mothers and fathers. On this scale are such items as: work out most quarrels with other children; take part in parents' conversations; choose to join (or not to join) Scouts or other organizations.

These findings allow the conclusion that for this group of parents, those mothers and fathers whose sons were academically successful, expected earlier mastery of common activities than did parents whose boys were nonperformers. This interpretation supports earlier research which shows a relationship between parental demands for certain types of behaviors and the achievement of their youngsters (Kessler, 1966).

2. An interesting finding pertained to the difference between successful and unsuccessful parents with respect to their values and training goals. This difference was assessed by reference to the frequency pattern of those 25 items selected as being most important as such were distributed across the six different scale categories. There was no real difference between groups on the Cognitive Facility Scale, all parents valued highly this type of activity. Conversely, items from the Physical Skills category were not selected by either group as being among the "more important." On the remaining four scales, however, differences were significant at $p < .01$. Successful parents chose items from the Autonomy category more frequently than did unsuccessful mothers and fathers; their mean number of choices from this scale was 13.8 as against 8.6 for the unsuccessful group. The reverse was true for items from the Wide Experience category. Unsuccessful mothers and fathers selected almost twice as many items from this scale as did the successful group. Since the Wide Experience scale includes such activities as "stay over night with a friend, go away to summer camp . . ., walk around town freely . . .," items which necessitate leaving home for a period of time, this finding provides substance for whimsical speculation. Further, the successful group emphasized Social Responsibility items to significantly greater extent, while their comparison parents chose Self-Care tasks as being the relatively "more important" for their son to master.

Thus, one finds an interesting difference in the *pattern* of values reflected by these category choices: successful parents stressed Social Responsibility, Autonomy, and Cognitive Facility. The unsuccessful group also valued Cognitive Facility but placed proportionately greater emphasis on activities designated as Self-Care or Wide Experience.

3. Closely connected to the above—and not unexpected—was the finding that parents of successful boys were in greater agreement ($p < .001$) about the particular scale items chosen as "more important"; that is, they also agreed far more on the *specific behaviors* which were most significant to them as parents. Although both sets of parents disagreed more than they agreed, successful mothers and fathers selected the same item 328 times, the unsuccessful mothers and fathers agreed only 250 times, suggesting that the successful parent had given greater thought to various childhood activities

and were in agreement about the relative significance of such. In short, they appeared to have a hierarchy of developmental tasks important for their child to experience. As example, should he be encouraged to have a paper route and, if so, at what age?

As previously noted, a secondary purpose of the research was investigation of parents' attitudes toward aspects of school thought to be particularly related to school success or failure. To this end, comments made by both groups of parents during the interview sessions were recorded and classified according to the following areas: (1) remarks about homework practices, (2) reasons given for son's academic success or failure, (3) major aspirations for child or concerns about son expressed during interview. From these, it was apparent that the successful parents took, for the most part, a democratic–authoritarian approach to child-rearing, held compatible views as to the role of a parent, and did not actively involve themselves in their son's homework. Representative examples from both groups are given below.

Successful Parents

"If we help one, we'd have to help all."

"It is his business and we hope he enjoys school. . . . We have tried to develop and maintain broad interests."

"No [help] but we check homework sometimes because if it is a theme he just can't be bothered with periods, etc. . . . sometimes I ask him to read to me so I'll know where he is reading."

"Only time he has trouble is when he is sick and he balks when lots [of homework] is sent home at once. . . . We tell him to break it up and do it in bits and pieces—not all at once."

"We always do things as a family and try to go places that are interesting to the children. At dinner, the children discuss their days with us."

Unsuccessful Parents

"[Son] hasn't done well lately because I've been working late and haven't been there to help."

Parents require their children to do an hour's schoolwork per day—regardless of whether or not they have homework of their own. They don't believe he could or would do work by himself.

"Won't do homework—not unless somebody stands over him." Father thinks he could and should!

Cause of 6th grade son's failure: "His second grade teacher."

"Our kids can't even object. We're hard grinders, know how to yell and scream just like the rest of the parents. . . . They want to talk at the table, but if they talk, they won't eat. . . . We don't have much to talk about anyway."

Implications of Research

A most practical outcome from this research was recognition of a pressing need for parent education. From parent interviews it was clear that those mothers and fathers heavily involved in their sons' homework were confused, as well as concerned, about what the

teacher expected both from them and from their child. Their misunderstanding was encouraged by the fact that there is often no uniform practice regarding homework among teachers by grade level. Overly concerned about their child's academic success, they made his work theirs, not recognizing that what appeared as giving help and support could, in reality, mean severe restriction of the boy's independent activity and prove a very real hindrance to his progress.

Another clear implication from this study is that children in trouble can be spotted early. The older elementary boys who formed the nucleus of the unsuccessful group had been in academic difficulty almost since they started school. The key grades were those in effort and study skills. We can and should get to these youngsters early in their school experience.

The results also underscore the importance of including fathers in discussions about their children. The fathers brought into the present research were unaware of the difficulties faced by their unsuccessful sons, either because their wives had not told them or because they failed to recognize that the problems were a little more serious than simply a matter of their sons being regular fellows. By contrast, the fathers of successful boys seemed to enjoy a much closer relationship with their sons (see Coopersmith, 1967, p. 109) and to be more keenly aware of what they were up to in school. Although scheduling was a problem, both groups of fathers welcomed the opportunity to talk with a member of the school counseling staff.

A challenging implication gleaned from findings is that, while the school does indeed have a child five or more hours per day, the youngster brings with him attitudes, behaviors, and expectations which influence the extent to which he will master successfully the school program. As Crandall and others have found (Crandall, Dewey, Katkovsky & Preston, 1964, p. 54),

> *. . . It is apparent to the careful observer that most children have developed, by the time they enter grade school, thoroughly consistent differences in the values they attach to intellectual and academic achievement, in their expectations of success in these activities, in the standards they use to judge their efforts, and in the methods and strategies they employ in an effort to gain achievement goals.*

These findings allow the conclusion that mothers and fathers whose boys were doing well in school not only expected appreciably *earlier* mastery of common tasks and activities, but also stressed different developmental *areas,* such as Autonomy, and agreed more as to the *particular* tasks and activities which they wished their sons to learn to do independently. From such, it appears that we may become too accustomed to thinking of underachievement solely in terms of socio-economic deprivations. This is indeed a major factor; we have Head Start and other programs for this reason. It should perhaps also be recognized that the child who simply comes to school curtailed in his efforts toward autonomy and mastery—to be independent within an age-appropriate framework—is unneces-

sarily handicapped by another form of deprivation—one that may not often enough be given adequate consideration.

References

Bentzen, F. Sex ratios in learning and behavior disorders. *American Journal of Orthopsychiatry,* 1963, **33**, 92–98.

Coopersmith, S. *The Antecedents of Self-Esteem.* San Francisco: W. H. Freeman, 1967.

Crandall, V., Dewey, R., Katkovsky, W., and Preston, A. Parents' attitudes and behaviors and grade school children's academic achievements. *Journal of Genetic Psychology,* 1964, **104**, 53–66.

Erikson, E. H. *Childhood and Society.* New York: W. W. Norton, 1950.

Gray, Audrey W. Independence concepts of parents concerning elementary school children. Unpublished doctoral dissertation, Claremont Graduate School, 1968.

Hobbs, N. Helping disturbed children: Psychological and ecological strategies. *American Psychologist,* 1966, **21**, 1105–1115.

Keith, R. A. The freedom to fail. In M. P. Douglass (Ed.), *30th Yearbook Claremont Reading Conference.* Claremont, Calif.: Claremont Graduate School and University Center, 1966, pp. 111–116.

Kessler, J. *Psychopathology of Childhood.* New Jersey: Prentice-Hall, 1966.

Levy, D. M. *Maternal Overprotection.* New York: Columbia University Press, 1943.

Shaw, M. C., and McCuen, J. T. The onset of academic underachievement in bright children. *Journal of Educational Psychology,* 1960, **51**, 103–108.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review,* 1959, **66**, 297–333.

2.3 *Children's Self-Concept and Kindergarten Achievement*

RICHARD J. OZEHOSKY AND EDWARD T. CLARK

Reprinted with minor abridgment from the *Journal of Psychology,* 1970, **75,** 185–192, with permission of the authors and the Journal Press, Inc. Richard J. Ozehosky and Edward T. Clark have both been school psychologists and are presently on the faculty of St. John's University, Jamaica, New York.

Ozehosky and Clark make a number of points relative to the relationship between children's self-concepts and their school performance. Their research shows that kindergarten teachers are able to perceive, with a fair amount of accuracy, the way in which children actually perceive themselves. These self-perceptions, in turn, are related to children's competence as measured by teacher ratings and reading readiness test scores. The study serves as an example of many which demonstrate the interaction of affective and cognitive factors in behavior.

A. Introduction

Many leading theorists have maintained that the self-concept plays a significant role in the education process (1, 7, 11, 12, 15). The contention is that when a child is accepted, approved, respected, and liked for what he is, he will have an opportunity to acquire an attitude of self-acceptance and respect for himself. With such an attitude he will have the freedom to venture forth into the school situation, and use his intelligence to its utmost capacity. A review of the literature reveals, however, that despite these claims little empirical evidence is available to support these contentions. Wylie (17) has noted that studies in the area of self-concept are filled with conflicting and confusing evidence. She ascribes the confusion to the great methodological difficulties inherent in the measurement of the self-concept, and to the bewildering array of hypotheses, measuring instruments, and research designs utilized.

Most of the research into the self-concept has been concerned with college students. McCandless (13, p. 179) states: "Compared with the number of adult studies, there are few studies of the self-concept in children. Self-concept studies of children younger than about fourth grade age are rare, since each child would have to be examined individually (because of reading and writing difficulties)."

Gordon and Combs (10) pointed out in their review of the literature relating the learner to self and perception, that while the literature contained many discussions of the origins and the development of the child's self, research was scant. They call attention to the pressing need for research both of a longitudinal nature and of normative studies at all development levels, theoretically based upon the development of the self-concept.

Dubin and Dubin (6) also point out the need for research in this area. They state, after a review of the literature, that from a methodological standpoint the empirical literature appears inadequate to support broad generalizations by virtue of scatter in theoretical interest and limited research technologies employed. They point out that the complicated processes relating self-concept to behavior have scarcely been studied.

The present study was concerned with ascertaining the relationship between children's self-concept and achievement in kindergarten. It attempted to answer the question: "Does the self-concept have functional utility at the kindergarten level?" It was predicted that girls would be significantly superior in achievement in kindergarten and in their perceptions of themselves than would boys.

B. Subjects and Procedures

The initial population in the present study consisted of 1042 children enrolled in the 37 kindergarten classes of a New York State public school district. At the conclusion of a three-session orientation program held in November, 19 teachers rated their pupils in terms of their judgment of the child's self-perception in the dimensions of competence-incompetence, good-bad, and ego

strength. The rating scale used by the teachers for this purpose was scored from zero to 11.

In May the teachers rerated the highest and lowest boys and girls in each class on the same rating scale utilized in November. The teachers' ratings of these 148 children proved to be quite stable—$N = .87$. Criterion groups were established by selecting the 25 highest boys and girls and the 25 lowest boys and girls in this sample.

Those children thus selected were individually administered the Quantified Self-Concept Inventory (16). This sentence completion instrument consists of a series of 40 stems based on the developmental areas used by Gesell and Ilg (9) and developmental tasks as listed by Havighurst (11). The child's responses to each item were coded, when relevant, as reflecting positive or negative self-perceptions. A global self-concept score for each child was derived by dividing the number of positive by the number of negative self-perceptions.

The children who had been identified for further study were also individually administered a nonverbal pictorial self-concept scale, developed by the authors (4)—namely, the U-Scale. This instrument was presented to the child as a game in which the child's task was to select (by pointing) the real "You" in each of 50 pairs of bipolar drawings. Each pair of drawings showed the male or female "You" figure in a positive or negative situation; e.g., "You" being rewarded or rebuked by his teacher. The 50 choices reflect such dimensions of self-concept as appearance, competence, and interpersonal relations. There are separate forms for boys and girls; the content is the same, only the sex of the "You"-figure and other protagonists are different: Each positive choice of the child was scored one point, with the maximum possible score being 50.

In the month of May all children were administered the Metropolitan Readiness Tests as part of the regular school testing program. This test provides scores for (*a*) Word Meaning, (*b*) Sentences, (*c*) Information, (*d*) Matching, (*e*) Numbers, (*f*) Copying, (*g*) Reading Readiness, (*h*) Number Readiness, and (*i*) Total Readiness.

In order that the kindergarten grading system might be adequately analyzed statistically, the four grading criteria utilized by teachers were given numerical weights. Satisfactory received a weight of 3, while Improvement Shown, Improvement Needed, and Unsatisfactory received weights of 2, 1, and zero, respectively. Thus it was possible to compute a Grade Point Average (G.P.A.) for every pupil, based upon the 26 report card categories.

C. Results

The question posed at the onset of the present investigation, "Does the self-concept have functional utility at the kindergarten level?" was answered positively. Table 1 supplies evidence indicating the utility of the various instruments employed in the present investigation in discriminating between the high and low criterion groups. It shows that the Metropolitan Readiness Tests were able to discriminate between high and low boys and girls at the .01 level of significance. Table 1 also indicates that

TABLE 1. Tests for Significance Between Mean Metropolitan Readiness, Mean Grade Point Average (G.P.A.), Mean U-Scale, and Mean Quantified Self-Concept Inventory Scores of High and Low Boys and Girls

Subjects	Mean	SD	*f*	*t*
		Mean Metropolitan Readiness Test Scores		
Boys				
High	77.00	9.33	2.65*	7.26**
Low	50.56	15.20		
Girls				
High	76.72	12.04	2.43*	5.43**
Low	52.00	18.77		
		Mean Grade Point Averages (G.P.A.) Scores		
Boys				
High	77.08	1.29	75.51**	6.14**
Low	62.64	11.45		
Girls				
High	77.24	1.30	92.11**	4.67**
Low	65.24	12.51		
		Mean U-Scale Scores		
Boys				
High	39.44	5.97	1.05	5.52**
Low	29.04	5.92		
Girls				
High	36.92	7.40	1.74	4.17**
Low	29.00	5.61		
		Mean Quantified Self-Concept Inventory Scores		
Boys				
High	3.72	3.35	3.85**	.61
Low	3.25	1.71		
Girls				
High	3.39	3.22	5.66**	1.15*
Low	3.00	1.35		

Note: A one-tailed *t*-test was used.
*Significant at the .05 level.
**Significant at the .01 level.

such was the case for Grade Point Average and that the U-Scale discriminated at the .01 level of significance between high and low boys and girls. However, the Quantified Self-Concept Inventory proved to be a poor discrimination between high and low boys and discriminated only at the .05 level of significance between high and low girls.

It was hypothesized that girls would be significantly superior in achievement in kindergarten and in their perceptions of themselves than would boys. Table 2[1] shows no significant differences on the

[1]Table 2 has been deleted in the interests of brevity.

Metropolitan Readiness Tests (M.R.T.) between high boys and girls, and low boys and girls and also shows no significant differences in Grade Point Average between high boys and girls and low boys and girls.

Table 3[2] indicates no significant differences in perception of themselves by high boys and girls measured by the Quantified Self-Concept Inventory (Q.S.C.I.). However, the Q.S.C.I. did show a significant difference of perception of self between low boys and low girls at the .01 level of significance in favor of low boys. Table 3 shows no significant differences in perception of themselves of high boys and girls and low boys and girls, as measured by the U-Scale.

Girls were not significantly superior in either achievement or in their self-perceptions. On the contrary, low boys as assessed by the Quantified Self-Concept Inventory perceived themselves significantly better, at the .01 level of significance, than did low girls. This significant difference in self-perception was not, however, supported by the U-Scale. No significant differences between boys and girls were found on the achievement measures.

D. Discussion

In order to fully comprehend the results of the present investigation, one must take into consideration the unusual nature of the sample studied. Two "criterion groups" were set up by the investigator, based upon teacher ratings of self-concept. These "criterion groups" consisted of the 50 highest self-concept (25 high boys and 25 high girls) and 50 lowest self-concept (25 low boys and 25 low girls) boys and girls. Some idea of the extreme dichotomy that evolved can be obtained when one considers that the "criterion groups" represent the highest and lowest self-concept children derived from an initial population of 1042. Extreme caution must therefore be exercised in the interpretation of data derived from the study of such a population.

Regarding the functional utility of the self-concept at the kindergarten level, it was demonstrated that at this age level self-concept is related to achievement. It was also shown that teacher judgments of children's self-concept at this early developmental level are predictive of kindergarten achievement, the correlation between teacher judgment and the Metropolitan Readiness Tests being a highly significant .70, significant at the .01 level.

It is of some importance to note that not only were teacher ratings of children's self-concept predictive of achievement at this level, but that teacher ratings of children's self-concept, at least at the extremes, was shown to be stable over a period of time. The teachers were asked to rerate 148 children, as to their self-concept, five months after their initial ratings of these children; a reliability coefficient of .87 was obtained.

The teacher ratings furthermore indicated a definite congruence between their perceptions of children's self-concepts and the child's self-concept, as measured by the U-Scale. The U-Scale measurement attested to the validity of the "criterion groups," which, of course, were based originally on teacher ratings

[2]Table 3 has been deleted in the interests of brevity.

of self-concept. The significant differences, at the .01 level, between high and low boys and girls supplied ample evidence as to the existence of a true dichotomy of self-concept.

It is interesting to note that the verbal instrument utilized to measure self-concept in the present study—namely, the Quantified Self-Concept Inventory—showed no such congruence between teacher ratings of self-concept and self-concept as actually measured by that instrument. It is quite apparent, therefore, that the nonverbal measure of self-concept, the U-Scale, and the verbal measure of self-concept—the Quantified Self-Concept Inventory—suggest contradictory findings. The former attests to the validity of the "criterion groups," while the latter invalidates the criterion grouping.

Subsequent analysis of the Quantified Self-Concept Inventory indicates that this instrument lacks validity when utilized on the present sample. A study of the item analysis of the Quantified Self-Concept Inventory indicates that the apparent significant differences in self-concept actually derive from the fact that 36 percent of the low girl responses, 29 percent of the high girl responses, 30 percent of the low boy responses, and 24 percent of the high boy responses could not even be scored. Thus the Q.S.C.I. was unable to adequately differentiate between the "criterion groups."

The supposition that girls would be significantly superior in achievement in kindergarten and in their perceptions of themselves than would boys was not confirmed by the results of the present study. These results contradict the findings that have been reported by other researchers, such as Dubin and Dubin (6), Davidson and Lang (5), and Fink (8), who based their conclusions on research conducted with older children.

Perhaps one explanation of the findings of the present study can be obtained by reporting Bledsoe's (2) study. He found significant differences in mean self-concepts of boys and girls at both the fourth grade and sixth grade levels, indicating that at these levels girls have greater self-esteem than boys. He hypothesized that the reason for this may be that the more frequent contacts with women teachers and with mothers (as compared with fathers) enables girls to develop a more satisfying self-concept at these developmental stages.

The findings reported in the present study seem clearly to confirm Bledsoe's hypothesis. It would appear that at the kindergarten level the factor of more female contact, contributing a derogatory effect on male self-concept and thus upon achievement, has not as yet begun to exhibit its effects. It can be said that this factor continues to operate and its derogatory effects on male self-concept and thus on achievement is measurable some years later in the developmental schema. This finding, if true, is of significant importance to educational and developmental psychology. It lends significant support to those who have been proposing that more male teachers be placed in the lower grades.

The results of the present investigation lent strong support to Brown's (3) cogent comments regarding the measurement of the self-concept at early developmental levels. He points out the considerable difficulties involved in reliably assessing the dimensions of self-concept among young children. He

comments that young children exhibit a generally limited ability to clearly verbalize complex self feelings and perceptions. Another major source of difficulty in the assessment of self-concept among young children, according to Brown, stems from the devices and procedures on which we have come to rely. We have all too often tended to rely on the downward revisions of techniques originally developed to measure self-concept among adults. This has resulted in the utilization of procedures that are often of limited use with young children due, says Brown, to their dependence on the ability of the subject to explore his feelings about himself in depth and to verbally report the essence of these self feelings to an adult.

Perkins and Shannon, based upon the results of their 1965 study (14), had suggested that techniques other than verbal self-report approaches might be suitable for assessing self-perceptions, and also that the pictorial identification techniques with children had a novelty which quickly aroused interest and attention. The present study bears out their contentions.

References

1. Allport, G. W. The ego in contemporary psychology. *Psychol. Rev.,* 1943, **50**, 451–470.
2. Bledsoe, J. C. Self-concept of children and their intelligence, achievements, interests and anxiety. *J. Indiv. Psychol.,* 1964, **20**, 55–58.
3. Brown, B. R. The assessment of self-concept among four-year-old Negro and white children: A comparative study using the Brown IDS Self-Concept Reference Test. Paper presented to the 37th annual meeting of the Eastern Psychological Association, New York City, April, 1966.
4. Clark, E. T., and Ozehosky, R. Preliminary manual for the U-Scale. Unpublished manual, Department of Psychology, St. John's University, New York, 1966.
5. Davidson, H., and Lang, G. Children's perceptions of their teachers' feelings toward them related to self-perception, school achievement, and behavior. *J. Exper. Educ.,* 1960, **29**, 107–118.
6. Dubin, R., and Dubin, E. Children's social perception: A review of research. *Child Develop.,* 1965, **36**, 809–838.
7. Erikson, E. H. *Childhood and Society.* New York: W. W. Norton, 1950.
8. Fink, M. B. Self-concept as it relates to academic underachievement. *Calif. J. Educ. Res.,* 1962, **13**, 57–62.
9. Gesell, A., and Ilg, F. *The Child From Five to Ten.* New York: Harper, 1946.
10. Gordon, I. J., and Combs, W., The learner: Self and perception. *Rev. Educ. Res.,* 1958, **18**, 433–444.
11. Havighurst, R. J. *Developmental Tasks and Education.* Chicago, Ill.: Univ. Chicago Press, 1948.
12. Jersild, A. T. *In Search of Self.* New York: Bureau of Publication, Teachers College, Columbia Univ., 1952.
13. McCandless, B. R. *Children and Adolescents.* New York: Holt, Rinehart, & Winston, 1961.
14. Perkins, C., and Shannon, B. Three techniques for obtaining self-perception in preadolescent boys. *J. Personal. & Soc. Psychol.,* 1965, **2**, 443–447.
15. Rogers, C. *Client-Centered Therapy.* New York: Houghton Mifflin, 1951.
16. Wattenberg, W., and Clifford, C. Relationships of the Self-Concept to Beginning Achievements in Reading. Detroit, Mich.: Wayne State Univ. Press, 1962.
17. Wylie, R. *The Self-Concept.* Lincoln, Neb.: Univ. Nebraska Press, 1961.

2.4 *A Study of Seventh Grade Children's Reading of Comic Books as Related to Certain Other Variables*

W. PAUL BLAKELY

Reprinted from the *Journal of Genetic Psychology,* 1958, **93,** 291–301, with permission of the author and The Journal Press, Inc. W. Paul Blakely is on the faculty of the School of Education, Drake University, Des Moines, Iowa.

Even experts in the field of education and mental health have been known to base their conclusions on "common sense," without bothering to put them to the test of research. The frequently heard statement that comic books incite children and youth to violence is an example of this. The present study, however, shows the desirability of putting "common sense" to the test, for it found no relationship between reading comic books and problem behavior.

A. Introduction

Speculation about alleged dangers as well as possible benefits of the ubiquitous comic books to their numerous youthful readers, has been widespread and vehement. There is less empirical data than the important status and serious implications of this problem would seem to warrant. Gunnar Dybwad, Executive Director of the Child Study Association of America, speaking before the United States Senate Subcommittee Investigating Juvenile Delinquency in 1954, declared that: "The absence of any definitive studies of the effects of comic reading on children's emotions and/or behavior has been a serious handicap to us as to everyone dealing with this problem." The Senate Subcommittee itself, in its Interim Report issued in 1955, noted "with some surprise that little attention has been paid by educational and welfare agencies to the potential dangers, as well as benefits, to children presented by the growth of the comic book industry."

The purpose of the present study has been to identify among a specific group of children, subgroups whose reading of comic books falls into various categories as to amount and as to type; then to measure as well as possible among these children certain variables upon which the reading of comic books has been alleged to have a bearing; and to determine whether the subgroups differ on the average with respect to these variables as measured.

B. Related Literature

It is rather easy to find literature in which this problem is discussed in terms of clinical or case-study material, or in terms of "expert opinion." It is more difficult, as already implied, to find reports of experimental data.

Dr. Frederic Wertham, in his status as a practicing psychiatrist, has commanded a wide audience with various pronouncements condemning comic books. Citing case material on children with reading difficulties whom he has found to be "addicted" to the reading of comic books, for example, he infers a casual connection (8).

Sperzel, in an experimental study involving use of comic books in the classroom, found no effects on reading ability (7). Luckiesh and Moss have cited statements of ophthalmologists that some children appear to suffer from eyestrain attributable to reading the comics, but have noted at the same time the lack of evidence available (2).

Witty has made inferences about comic reading and educational achievement in a group of Evanston-Chicago-Milwaukee children he studied, reporting that a subgroup who read the most comic books received almost the same average school marks as one who read the fewest (9). Heisler (5) studied 600 elementary and junior high school children and for each school grade level, made a number of comparisons between a group who read no comic books and one of equal or nearly equal size who read the most. Evidences of differences in school achievement, adjustment, and ownership of library-type books were neither consistent nor of a generally recognized significance level.

Concerning adjustment and behavior, Dr. Wertham (8) says that his clinical data obtained by a number of standard techniques show factors in the psycho-

logical patterns of children which require other than conventional Freudian explanations, and correspond to material in comic books. He cites as an example, an injury-to-the-eye motif.

Josette Frank (4) in 1948 reported a symposium of psychiatric and psychological opinion, concluding that there was agreement that "the radio programs, movies and comics do not in themselves *create* fears, but for certain children and under various conditions, do precipitate or stimulate anxieties lying beneath the surface ready to be awakened."

Wolfe and Fiske studied a relatively small group of children intensively and found certain unhealthy characteristics associated with reading the comic books as "fans," that is, with excessive intensity (10).

Hoult (6) studied 235 boys and girls, ages 10 through 17, who had been arrested for juvenile delinquency, in comparison with a matched group of controls. He categorized comic books as harmful, questionable, and harmless according to their general content (e.g., crime and gangsterism, animated animal cartoons), and found that his delinquent subjects reported reading more titles in the first two categories than the nondelinquents, but no more in the third.

C. The Present Study: Population and Method

The population chosen for the present study consisted of 12 class sections of seventh grade children in two public junior high schools in an Iowa city of about 100,000 population. Complete data were obtained for 323 children, except that one of the coöperating schools withheld a group of children from the achievement testing program so that achievement test scores were available on only 281.

The city is one which thrives on industry and trade. The schools in the study draw from populations which are not atypical in the city. An unselected list of parents' occupations as reported by the children in the study follows: printer, super-market manager, tool-grinder, worker in sales department of radio-manufacturing company, truck-driver, dock foreman, traveling salesman, crane operator, dishwasher (mother), painter, electrician, design engineer, roofer, and college instructor.

As a measure of comic book reading, three questionnaires were constructed, each labeled *Junior High Interest Questionnaire,* and administered at intervals of about one month beginning early in September, 1956. Among the questions on each, the child was asked to list the titles of all comic books he had read "within the past week," and in any case where he had read the same title more than once, to put the number of times just following the title. The question about comic book reading, in each case, was placed among others about various interests and activities in an effort to avoid the possibility of putting the responding child "on his guard"; that the study was particularly concerned with comic book reading was not divulged.

From the questionnaires was computed for each child his total comic book *frequency*—that is, the total number of titles reported read, multiplied by the number of times for each. (For example, having read *Donald Duck* 2

times and *Tarzan* once would constitute a *frequency* of 3.)

For each child was also computed a *comic book type score.* Each title which appeared in the most recent classification of the Cincinnati Committee on Evaluation of Comic Books was considered accordingly, as rating "No Objection," "Some Objection," or "Objectionable." For titles not on the most recent list, earlier lists were consulted. Some titles found on none of these lists were classified by the investigator, applying as best he could the criteria set forth by the Committee in its 1956 leaflet. It was impossible to classify 7.31 percent of the reported frequency, which were ambiguous or could not be identified; these titles were ignored in computing the comic book type score. The complete list of titles reported, with their classifications, appears in the writer's doctoral dissertation, which is in the library of the State University of Iowa, and on file with University Microfilms, Ann Arbor, Michigan.

The comic book type score for each child was computed by the formula

$$T = \frac{(N \times 2) + (S \times 1) + (B \times 0)}{N + S + B} \times 10$$

where T = type score; N = frequency of "No Objection" titles; S = frequency of "Some Objection" titles; and B = frequency of "Objectionable" titles.

As a measure of "legitimate" reading, each questionnaire also asked the child to list all books of the type obtainable at a library, which he had read "within the last month," and after each, to specify where he had obtained it—public library, school library, home, etc.

As an intelligence test as independent of reading as possible, The Lorge Thorndike Intelligence Test, Non-Verbal Series, Level 4, Form A was administered.

The SRA Junior Inventory, Form S was chosen as a measure of various phases of adjustment. The checking of problems on the Inventory as "big" or "middle-sized" has been taken as indicative of the child's concern about his adjustment in the respective area. Checking a problem as "little" or "not a problem" might, of course, indicate conscious or unconscious reluctance to admit the problem, as well as the nonexistence of the problem for the child concerned. Therefore, the scores derived by counting the number of checks in "big" and "middle-sized" boxes must be taken as indicative of the child's *perception* (quantitatively) of his problems. For each child, the total score on the inventory was thus computed, as well as the subscores "About Me and My School," "About Me and My Home," "My Health," "About Myself," "Getting Along With Other People," and "Things in General"; and a subscore for a category of items selected throughout the inventory taken to be indicative of "personal adjustment."

As a measure of behavior and behavior-related characteristics observable to others, teachers of the children filled out the Haggerty-Olson-Wickman Behavior Rating Schedules, *A* and *B.* Schedule *A* consists of a list of behavior problems in their order of frequency as reported for a group of elementary school children, with statistical weights assigned on the basis of seriousness and frequency. In Schedule *B* are five-point graphic rating scales for traits

studied in relation to the behavior problems of Schedule *A* among the standardizing group; thus a high score on Schedule B is taken to be suggestive of behavior-problem proneness.

The Stanford Achievement Test, Advanced Battery, Form *J,* was administered during the school year as part of the regular program of the cooperating schools. Grade norm scores for the children in the study were taken as follows: reading (average of Paragraph Meaning and Sentence Meaning); average of Language and Spelling; and general school achievement status (average of Language, Spelling, Arithmetic Reasoning, Arithmetic Computation, and Social Studies).

The intelligence test, the *SRA* Junior Inventory, and Questionnaires 1 and 2 were administered by the investigator during school periods allotted him for the purpose. Questionnaire 3 and the Standard Achievement Test were administered, and the ratings on the Haggerty-Olson-Wickman Schedules made, by coöperating school personnel.

D. Parent Interviews

Hoping to be able to make an assessment of the truth of the children's answers to the questionnaires, the investigator visited the homes of 35 of them. The 35 were initially chosen by taking every tenth name from an alphabetized list of all children who responded to the first questionnaire. When an interview could not be completed after a second call at a home (because no one was found at home), the next name in order was substituted.

The parents were asked questions about their children's comic book and library book reading, along with others about television viewing and other activities. It soon became apparent that an estimate of their children's weekly reading of comic books is difficult for many parents to make. Likewise, much of the library book reading was found to take place in the school, and the parents were unable to report it very accurately. A product-moment correlation coefficient of .42 was obtained, for the 35 cases, between the average weekly frequency of comic book reading computed from each child's three questionnaires, and the parents' estimates of "number of comic books read each week." A coefficient of .43 was obtained between the average monthly number of library books reported read by the children, and a parent estimate of library books obtained monthly, on the average, from the public library plus "other sources."

Many of the parents interviewed obviously felt it reflected on their child-rearing to admit their children were readers of comic books. This, plus the fact that the children were asked actually to list titles of comic books and library books read, is taken as justifying the conclusion that there is probably greater error in the parents' estimate than in the figure derived from the children's reports.

E. Results of the Study

1. Comparisons based on comic book frequency groups

The children participating in the study were divided into subgroups according

TABLE 1. The Subgroups According to Total Reported Comic Book Frequency for Three Non-Consecutive Weeks

Subgroups	*N* [a]	Range	Mean	*SD*
Boys				
I. Non-readers	43 (30)	0	0	0
II. Light readers	35 (32)	1–6	3.26	1.63
III. Moderate readers	55 (46)	7–21	12.95	4.23
IV. Heavy readers	39 (34)	22 or more	44.18	27.76
Total	172 (142)		14.82	21.44
Girls				
I. Non-readers	33 (27)	0	0	0
II. Light readers	46 (44)	1–6	3.48	1.53
III. Moderate readers	54 (51)	7–21	12.07	3.96
IV. Heavy readers	18 (17)	22 or more	39.67	27.44
Total	151 (139)		10.11	15.41
Total				
I. Non-readers	76 (57)	0	0	0
II. Light readers	81 (76)	1–6	3.38	1.58
III. Moderate readers	109 (97)	7–21	12.51	4.12
IV. Heavy readers	57 (51)	22 or more	42.75	27.75
Total	323 (281)		12.62	19.01

[a] Figures in parentheses are reduced *N*'s used for achievement test comparisons.

to total reported comic book frequency, for purposes of a series of comparisons. An attempt was made to keep subgroups of maximum size possible while at the same time maintaining categories descriptive as to frequency of comic book reading. Table 1 shows the subgroupings used; it will be noted, for example, that "Light Readers" reported reading frequency of from one to six comic books during the three weeks covered by the questionnaires, or an average frequency of not more than two per week.

Comparisons among the subgroups so derived, with respect to the other measures, which are regarded as dependent variables, were made by the simple analysis of variance, using .05 coefficient of risk. The comparisons are shown in Table 2.

Where the obtained *F* meets the chosen criterion of significance, the *t*-test has been applied between the various pairs of means. The following differences in reported library-book reading were thus found to be significant: Subgroups III over I and IV over I (.01 level); and IV over II (.05 level).

When the same set of analyses of variance were applied to the sexes separately, an *F* of 4.83 (significant at .01 level, $df = 3,168$) was obtained for boys alone in reported library-book reading; and significant *t*'s were found in testing the following differences: Subgroups III over I, IV over II, and IV over III (.05 level); and IV over I (.01 level). Thus for boys, and for the sexes combined, reported readership of library-type books tends to increase as does frequency of comic book reading. The

subgroup means for library book readership for girls clearly followed the same pattern, but not to the chosen level of significance.

With the exception noted, the comparisons made among subgroups who differ in frequency of comic book reading show no corresponding, average differences in the measures applied.

2. Comparisons based on comic book type score groups

A second arrangement of subgroups has been made according to comic book type scores, as shown in Table 3. Again, factors of both maximum subgroup size and descriptiveness of category were considered. However, the range of type scores in Subgroup IV is necessarily wider than might be desirable, because of the considerable negative-skewedness of the type-score distribution. It will be noted that Subgroup I, "Non-readers," is the same subgroup so-used in the comparison based on frequency.

Results of the simple analysis of variance as applied to this arrangement of subgroups, with respect to the various measures applied, may be observed in Table 4. Again, the lack of significant differences is much greater than their presence. It will be noted that the tendency of comic book readers to be greater readers of library books, by

TABLE 2. Comparisons Among Comic Book Frequency Subgroups (Sexes Combined)

Measure	Subgroup Means I	II	III	IV	*df*	*F*
Lorge-Th.						
Non-V IQ	99.53	102.73	102.81	102.40	3,319	1.55
Library Book						
Rdg. (3 mo.)	4.00	5.14	5.73	6.68	3,319	5.12**
SRA Jr. Inv.						
Home	2.20	2.31	2.59	3.26	3,319	1.30
People	1.46	2.48	2.28	3.00	3,319	1.91
School	8.50	7.67	7.89	8.35	3,319	0.28
Self	2.50	3.24	3.28	2.95	3,319	0.62
Th. Gen.	4.59	4.98	5.31	6.18	3,319	0.99
Health	1.93	2.22	2.44	2.75	3,319	0.86
Pers. Adj.	2.45	2.91	2.89	3.07	3,319	0.28
Total	21.18	22.89	23.80	26.49	3,319	0.65
Hag-Ol-Wick						
Sch. A	11.42	7.52	11.87	13.95	3,319	1.30
Sch. B	66.34	65.35	67.60	67.07	3,319	0.29
St. Achvt.						
Reading	9.01	8.95	8.99	9.14	3,277	0.10
Lang-Sp	8.52	8.56	8.31	8.32	3,277	0.37
5 Subtests	8.47	8.45	8.34	8.28	3,277	0.27

**Significant at .01 level.

TABLE 3. The Subgroups According to Comic Book Type Scores [a]

Subgroup	*N* [b]	Range	Mean	*SD*
Boys				
I. Non-readers	43 (30)	—	—	—
II. All "No Obj."	50 (44)	(All have scores of 20)		
III. Few "S.O.," "Obj."	46 (37)	16–19	17.67	1.06
IV. Most "S.O.," "Obj."	33 (31)	0–15	11.15	4.34
Total readers	129 (112)	0–20	16.91	4.19
Girls				
I. Non-readers	33 (27)	—	—	—
II. All "No Obj."	70 (66)	(All have scores of 20)		
III. Few "S.O.," "Obj."	30 (29)	16–19	17.50	1.06
IV. Most "S.O.," "Obj."	18 (17)	0–15	12.94	2.63
Total readers	118 (112)	0–20	18.29	2.75
Total				
I. Non-readers	76 (57)	—	—	—
II. All "No Obj."	120 (110)	(All have scores of 20)		
III. Few "S.O.," "Obj."	76 (66)	16–19	17.61	1.07
IV. Most "S.O.," "Obj."	51 (48)	0–15	11.98	3.92
Total readers	247 (224)	0–20	17.57	3.64

[a] For derivation, see text.
[b] Figures in parentheses are reduced *N*'s used for achievement test comparisons.

their own report, at least, persists without respect to the "type" of comic books read.

When the series of analyses were applied to the type-score subgroups, sexes separately, one further difference emerged. The girls in Subgroup III score higher on problems related to "Things in General" than do those in Subgroup II (see Table 5). This category of items in *SRA* Junior Inventory, Form *S*, is rather miscellaneous; the one outstanding characteristic seems to be a lack of skills or "know-how": "I need to learn how to use the library"; "I want to learn how to dance"; "I wish I could take music lessons"; "I wonder what my real abilities are." It is probably not justified to invest too much concern in this one significant difference which appears among so many possible ones (maybe this is an error of rejection which the .05 coefficient of risk implies will occur 5 per cent of the time in the long run). However, it should be noted that girls in type-score Subgroups III or IV, or both, also have higher observed means than those in Subgroups I and/or II (not at the chosen level of significance) in *SRA* total, "People," "Self," "Health," and "Personal Adjustment," and in the Haggerty-Olson-Wickman schedules.

F. Conclusions

Within the limitations inhering in the method and population of the study, the following conclusions have been drawn:

TABLE 4. Comparisons Among Comic Book Type Score Subgroups (Sexes Combined)

Measure	Subgroup Means I	II	III	IV	*df*	*F*
Lorge-Th.						
Non-V. *IQ*	99.53	103.39	101.54	103.00	3,319	2.00
Library Book						
Rdg. (3 mo.)	4.00	5.73	5.71	5.90	3,319	3.51*
SRA Jr. Inv.						
Home	2.20	2.33	3.11	2.73	3,319	1.20
People	1.46	2.08	2.79	3.14	3,319	2.59
School	8.50	7.53	8.84	7.47	3,319	0.88
Self	2.50	3.08	3.51	2.96	3,319	0.76
Th. Gen.	4.59	4.58	6.45	5.77	3,319	2.38
Health	1.93	2.32	2.57	2.55	3,319	0.65
Pers. Adj.	2.45	2.63	3.18	3.29	3,319	0.66
Total	21.18	21.93	27.26	24.61	3,319	1.26
Hag-Ol-Wick						
Sch. A	11.42	9.83	12.12	11.64	3,319	0.24
Sch. B	66.34	64.81	66.59	72.55	3,319	2.20
St. Achvt.						
Reading	9.01	8.89	9.15	9.09	3,277	0.31
Lang-Sp	8.52	8.40	8.63	7.94	3,277	1.51
5 Subtests	8.47	8.37	8.43	8.24	3,277	0.27

*Significant at .05 level.

TABLE 5. Comparisons Among Girls' Comic Book Type Score Subgroups, SRA *Junior Inventory, "Things in General"*

Sub-groups	Means	Differences I	II	III
I	5.36			
II	4.30	1.06		
III	7.57	2.21	3.27**	
IV	6.50	1.14	2.20	1.07

F ($df = 3, 147$) = 2.80.*
*Significant at .05 level.
**Significant at .01 level.

1. Seventh grade children who differ in comic book reading, either as to type (as defined) or as to frequency, do not differ correspondingly in tested reading ability.
2. Such children, so grouped, do not differ correspondingly in achievement in language and spelling as measured by a standardized test.
3. Such children, so grouped, do not differ correspondingly in general school achievement as measured by an average of five subtests of a standardized test.
4. Such children, so grouped, do not differ correspondingly in behavior problem status, or in traits related to behavior problem status, as observed and rated by their teacher.
5. Such children, so grouped, are not significantly different from each

other, correspondingly, in intelligence.

6. Seventh grade children who read comic books read more library books than do those who read no comic books. The difference tends to be in proportion to the frequency of comic book reading, and to exist for all "qualities" of comic book reading as defined by the Cincinnati Committee. Although the difference is not significant in the case of girls alone, their subgroup means clearly follow the same pattern as those for boys and they appear to contribute to the significance level for the combined sexes.
7. Seventh grade boys who differ in comic book reading, either as to type (as defined) or as to frequency, do not differ correspondingly in reported recognition of their own personal problems.
8. Seventh grade girls who read some comic books of the type regarded by the Cincinnati Committee as rating "Some Objection" or "Objectionable" are more troubled than are those who read entirely comic books of the "No Objection" type, by some personal problems. The difference is statistically significant in the *SRA* Inventory subdivision, "Things in General," which seems in general concerned with need for knowledge, skills, and experience. There is a suggestion, not otherwise statistically substantiated, that girls whose reading of comic books falls into these "objectionable" categories may be a somewhat less well-adjusted group than others. The point is highly inconclusive.
10. The data as analyzed provide only very meager evidence (and this in the case of girls only) that the type of criteria applied in classifying comic books according to type or "quality" identify comic books the reading of which has observable, immediate concomitants in the variables studied for seventh grade children.

Finally, it may be well to emphasize that this study does not purport to have disproved the existence of all the ills allegedly attendant upon the reading of comic books. It is a serious matter to endanger the morals, the mental health, or the normal progress of development of children. It is also a serious matter to interfere with popular access to the media of communication—certainly too serious to be undertaken casually. Those who would do either of these two things ask us to engage in a calculated risk of large proportions. Upon those proposing either type of risk it would seem incumbent to marshall justifying evidence. The results of this study for the most part *fail to support* curtailment of children's access to comic books.

References

1. Educational Digest. Comic books and juvenile delinquency. *Educ. Dig.,* 1955, **21**, 8–10.
2. Deboer, J. J. Using modern channels of communication: Magazines. *Elem. Eng.,* 1950, **27**, 107–125.
3. Dybwad, G. Statement to the Senate Subcommittee. *Child Stud.,* 1954, **31**, 38–41.
4. Frank, J. Chills and thrills in radio, movies, and comics. *Child Stud.,* 1948, **25**, 42–48.

5. Heisler, F. A comparison of comic book and non-comic book readers of the elementary school. *J. Educ. Res.,* 1947, **40**, 458–464.
6. Hoult, T. F. Comic books and juvenile delinquency. *Sociol. & Soc. Res.,* 1949, **33**, 279–284.
7. Sperzel, E. Z. The effect of comic books on vocabulary growth and reading comprehension. *Elemen. Eng.,* 1948, **25**, 109–113.
8. Wertham, F. *Seduction of the Innocent.* New York: Rinehart, 1954.
9. Witty, P. Reading the comics: A comparative study. *J. Exper. Educ.,* 1941, **10**, 105–109.
10. Wolfe, K. M., and Fiske, M. The children talk about comics. In (Lazarsfeld, P. F., & Stanton, F. N., Eds.), *Communications Research 1948–1949.* New York: Harper, 1949.

2.5 *Where Are the Siblings? A Re-evaluation of the Relationship Between Birth Order and College Attendance*

WILLIAM T. SMELSER
AND
LOUIS H. STEWART

Reprinted from *Sociometry,* 1968, **31,** 294–303, with permission of the authors and the American Sociological Association, Inc. William T. Smelser is a member of the faculty of the University of California at Berkeley, and Louis H. Stewart is at San Francisco State College. Both are clinical psychologists.

The foregoing articles in this section all point to a close relationship between the behavior of students and their early childhood experiences. The effect of parental attitudes and expectations is, as we have noted, particularly significant. Also important are the roles and position held by a child during these early years. The effect of these variables is perhaps best seen in research dealing with birth order; much research shows that first-born children are "success-oriented" and are more likely to go on to college than are those who were later born in their families. The present study shows that if the next youngest sibling is of the opposite sex, the efforts of the oldest appear to be enhanced and facilitated. Birth order is not the most important variable with respect to understanding the behavior of a learner, but the research does show how such factors as role, social position, and social setting interact to affect the behavior of children, youth, and adults alike.

In a review of studies of the relationship between birth order and college attendance (Altus, 1966) it has been reported that first-born children are more highly represented in college populations than are the later-born children. This relationship has been found to be consistent for both sexes and for a wide range of family sizes (Altus, 1965; Schachter, 1963). However, in none of the studies reported is cognizance taken of the fact that the samples do not contain siblings of the subjects. The first-borns of these samples come from one set of families and the intermediate and last-borns come from other completely independent sets of families.

The fact that *total* families are not represented may limit the generalizability of the findings, since factors other than birth order *per se,* could account for the higher representation of first-borns in college. One such factor is socioeconomic differences between families with first-borns in the college population and families with later-borns. A second factor is the frequency distribution of various family constellations in the college student population. This may be illustrated by the two-child families which are comprised of four possible ordinal position and sex of sibling constellations: boy–boy, girl–girl, boy–girl, girl–boy. If there should be any appreciable differential frequency of occurrence of these family constellations in the general population, then it would follow that various differential frequencies of first- and last-born males and females could well occur in the samples reported, opening up a possible source of bias. Another kind of limitation in these studies lies in the lack of a measure of *terminal* education, so that generalization to samples with either less than college education or post graduate education is not possible.

What is necessary to clarify the relationship of birth order to college education is: (1) a sample of complete families (i.e., *all* the siblings) of all family constellations in order to permit intra-family analysis of the relationship, where the true test of any sibling order effect must lie, as well as an analysis of the incidence of family constellations represented in college, (2) measurement of the socioeconomic status of the parents and (3) terminal education of all children. Data permitting such an analysis of the intra-family relationships of birth order and education is found in a recent follow-up study of a longitudinal sample of the Institute of Human Development, University of California, Berkeley, where the educational attainment of the subjects, their spouses, and the siblings of both subjects and spouses was obtained.

Samples and Method

The basic sample of this study consists of families ($N = 155$) of the Guidance Study (GS) and families ($N = 122$) of the spouses of GS subjects. The GS is a longitudinal study and description of the original sample is found in an earlier publication (Macfarlane, 1938). This sample is predominantly middle class, urban, and white. In 1928, the occupational composition of the sample was 27 per cent professional and executive, 41 per cent white collar, 20 per

cent skilled trade, 10 per cent semi-skilled trade, and 2 per cent unskilled labor.

Measures of the dependent variable, years of schooling, were obtained from the subjects as a part of the adult interviews conducted from 1957 through 1960. The means of age and education of the male and female subjects, spouses and siblings are found in Table 1. The years of schooling for the siblings of the Guidance subjects, as well as the siblings of the spouses, were obtained from the subjects and their spouses, respectively. Only subjects whose years of formal education were known were included in the study, and subjects from families with five or more children were not included, for there were too few of these larger families.

The possibility that there might be systematical differences due to a bias in *report* of education for the self (informant) and the siblings was investigated by comparing the years of schooling reported by the informant with the years of schooling reported for his (or her) siblings. The mean differences are clearly a joint function of both the sex of the informant and the sex of the siblings. Males report more schooling than their sisters, females report less schooling than their brothers, while reported differences between self and siblings of the same sex are negligible. Each of the four comparisons between self and sibling are statistically insignificant (Wilcoxon matched-pairs signed-ranks test). Of the 300 comparisons between informant and sibling, 39 per cent showed the informant reporting more education; 27 per cent reported equal education; and 34 per cent reported the sibling with more education. There is a slight and statistically insignificant tendency for the informant to report more schooling for the self than for the sibling(s). Consequently, the years of education of the siblings is not considered to be a significantly biased report and will be employed in subsequent analyses of intra-family differences.

Results

The data of the GS will first be presented in the format of previous research which does not attend to the fact that the subjects in different birth

TABLE 1. Age and Education: Sample Means and Standard Deviations

		Females			Males		
		Subjects	Spouses	Sisters	Subjects	Spouses	Brothers
	N	84	58	164	71	64	175
Age (1960)	M	31.6	30.2	32.8	31.5	34.7	33.1
	SD	.5	2.5	7.7	.5	3.8	8.0
Education	M	14.1	13.8	13.6	15.0	15.4	14.5
	SD	2.3	2.0	2.1	2.2	2.6	2.6

order positions come from different families and from different types of family constellations. Tables 2 and 3 present, for females and males respectively, the relationship between years of schooling and ordinal position. Designation of years of schooling in Tables 2, 3, and 4 is as follows: 17+, one or more years in graduate school; 16, graduation from college; 13–15, one to three years of college (this includes three subjects with four years of college without a degree); and 12, high school or less. (The number of four-child and larger families is too small to warrant inclusion in Tables 2 and 3.) In both the two and three-child families, it is clear that, for both sexes, first-borns graduate from college and attend graduate school in greater proportions than do last-borns. A comparison between *all* first-born (excluding only children) with all last-born shows that, for females, 49 per cent of first-borns graduated from college as compared with 30 per cent of last-borns, while for males, 63 per cent of all first-born graduated as compared with 51 per cent of all last-born males. These findings are comparable to those reported for attendance at college in previous studies. However, the presentation of the data in Tables 2 and 3 does not account for the critical factor of family constellation.

Analysis by Family Constellations

When the data of Tables 2 and 3 for two-child families only are rearranged to show the distribution of terminal edu-

TABLE 2. Distribution of Education for Females

Birth Order	Years of Schooling	Family Size 1	2	3	Total
1	17+	1 (3)*	7 (13)	3 (10)	11 (10)
	16	6 (18)	18 (35)	12 (40)	36 (31)
	13–15	9 (27)	18 (35)	7 (23)	34 (29)
	12–	17 (51)	9 (17)	8 (27)	34 (29)
		33	52	30	115
2			1 (2)	4 (12)	5 (6)
			15 (28)	13 (39)	28 (32)
			18 (34)	7 (21)	25 (29)
			19 (36)	9 (27)	28 (32)
			53	33	86
3			1 (3)		1 (3)
			9 (26)		9 (26)
			10 (28)		10 (28)
			15 (43)		10 (43)
			35		30

*Percentages in Parentheses.

cation as a function of *both* ordinal position *and* family constellation, we find somewhat different results (Table 4). Inspection of both the all female (FF) and all male (MM) two-child family distribution shows that there is no consistent difference in educational achievement favoring the first-born.

TABLE 3. Distribution of Education for Males

Birth Order	Years of Schooling	Family Size 1	2	3	Total
1	17+	7 (27)*	16 (30)	10 (27)	33 (28)
	16	7 (27)	20 (37)	11 (30)	38 (32)
	13–15	5 (19)	8 (15)	7 (19)	20 (17)
	12–	7 (27)	10 (18)	9 (24)	26 (22)
		26	54	37	117
2			10 (19)	11 (30)	21 (24)
			19 (36)	9 (25)	28 (31)
			10 (19)	10 (28)	20 (22)
			14 (26)	6 (17)	20 (22)
			53	36	89
3				7 (20)	7 (20)
				10 (28)	10 (28)
				10 (28)	10 (28)
				8 (23)	8 (23)
				35	35

*Percentages in Parentheses.

*TABLE 4. Education by Ordinal Position and Sex of Sibling in Two Child Families **

	All Female		All Male		Sister then Brother		Brother then Sister	
Education	1st	2nd	1st	2nd	1st	2nd	1st	2nd
17+	4	4	22	30	20	10	35	0
16	27	32	43	39	40	33	32	26
13–15	41	32	13	9	30	27	16	35
12–	27	31	21	22	10	29	16	38
N	22	22	23	23	30	30	31	31
M	14.2	13.9	15.1	15.5	15.3	14.3	15.6	13.5
SD	1.9	2.0	2.4	2.2	1.8	2.0	2.0	1.7

*Except for N, M, and SD, all figures are percentages.

However, by contrast, in sister then brother (FM) and brother then sister (MF) families, there is a marked tendency for the first-born child to attend college, to graduate from college and to go to graduate school more often than the last-born sibling. Ninety per cent of these first-born sisters attended college, while only 61 per cent of their younger brothers attended college; 84 per cent of first-born brothers attended college as compared with 62 per cent of their younger last-born sisters. Application of the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) to the male then female families (MF) reveals that the males achieve significantly more schooling than their younger sisters at the .001 level (two-tailed test). The Wilcoxon test also reveals, that, in the female then male families (FM) females achieve significantly more schooling than their younger brothers at the .05 level (two-tailed test). The Wilcoxon test reveals *no* significant differences in amount of schooling between males in two-child, all-male families or between females in two-child, all-female families.

The size of our sample does not permit a similar detailed analysis of families of size larger than two. However, some contributing evidence has been obtained from the following analysis of three and four-child families. A comparison has been made between subjects with sibling of opposite sex following and with sibling of the opposite sex preceding. These data are summarized in Table 5. For males, it is apparent that the educational achievement of males followed by a sister is higher than for males preceded by a sister; likewise, educational achievement for females followed by a brother is consistently higher than for females following a brother. For *both* males and females, then, more years of schooling is a function of preceding a sibling of the opposite sex, while fewer years of schooling is a function of following a sibling of the opposite sex.

Considering the data in Table 4, several inter-family comparisons are of obvious interest. First, the last-borns from the all-male families (MM) attain *more* schooling than do last-born males from the sister then brother (FM) families. Second, the last-born women in the MF families attain *less* schooling when compared with the last-borns in the FF families. However, in order to determine whether such *inter*-family differences

TABLE 5. Education in Three-Four Child Families as A Function of Sex of Adjacent Sibling *

Education	*Male* then Sister	*Male* following Sister	*Female* then Brother	*Female* following Brother
17+	28	13	12	3
13–16	61	53	44	26
12–	22	31	38	50
N	41	36	52	42

*Except for sample sizes, all figures are percentages.

are a function of the sex of the sibling it is necessary to test for possible socioeconomic differences between the families. A four-group, one-way non-parametric analysis of variance (Kruskal-Wallis H test) revealed no significant differences between the mothers; the same analysis also revealed no significant differences between the fathers. These insignificant differences make the above inter-family comparisons more tenable.

Frequency of Family Constellations

In addition to the first-born's greater amount of education in *both* combinations of cross-sex two-child families, there are more cross-sex two-child families than same sex two-child families. Of the total of the 106 two-child families in Table 4, 57 per cent are cross-sex families, while 43 per cent are same sex families. This finding is corroborated by the data from two other independent longitudinal samples of the Institute of Human Development research program (Jones, 1939; Bayley, 1966); of 83 two-child families, 54 per cent are cross-sex families as compared with 46 per cent same-sex families. Thus, not only is the educational discrepancy between first and last-born greatest in the boy-girl (MF) and girl-boy (FM) families, but it would appear there is a *greater proportion* of this sex composition and sequence family in the two-child families.

Discussion

Our results suggest that the widely reported finding that more first-borns attend college is not a phenomenon of birth order *per se,* but is instead an effect, first of all, of some interaction of birth order, sex of child, and sex of sibling; secondly for the two-child families, at least, the greater frequency of cross-sex two-child families. Both of these: the higher educational achievement of first-borns in mixed sex families and the overrepresentation of just such families in the population have been overlooked in previous investigations.

Many authors have speculated on the differential treatment of the children as a function of the sex and birth order of their offspring (e.g., Aberle & Naegle, 1952); others focused on the mutual impact of the siblings (Brim, 1958). Our findings raise several questions concerning the interpersonal influence between siblings and its relationship to differential educational achievement. Does the older sister in the two-child family act as an "intellectual depressant" for her younger brother while he in turn acts as an "intellectual stimulant" for her? On the other hand, in the brother then sister family, it appears that, while the younger sister may act as a stimulant for her older brother, he in turn does not act as an intellectual depressant for her. Her schooling is comparable with the schooling of the last-born sister in the all-female family, while the older brother attains more schooling than the first-born boy in an all-male family.

It would be of considerable interest if there were some independent evidence which supported the hypothesis of an interaction between birth order, sex of child and sex of the sibling which contributed to the outcome of more school-

ing for the first-born under conditions where the first-born was followed by a sibling of the opposite sex. Some such evidence does seem to exist in the research of Helen Koch (1955) in which comparisons on a variety of personality and behavioral traits have been made among children of two-child families. Her data seem to provide some highly suggestive support for our findings. Most significant is the fact that in Brim's reanalysis of Koch's data it was found that in a majority of traits the children with siblings of the opposite sex tend to differ from children with same sex sibs. In addition, Brim's analysis indicates that the one trait on which *both* first-born boys and girls with sibs of the opposite sex were rated higher than were first-borns with siblings of same sex, was, curiously enough, curiosity.

Further evidence in support for our second major finding, viz., the greater proportion of cross sex-two-child families, is found in a study of 905 married couples by Westoff, et al. (1963). Couples with two children of the same sex (MM or FF) reported a wish to have more additional children than did couples with a child of each sex (MF or FM). Furthermore, Westoff's follow-up study found that *in fact* couples with children of the same sex actually went on to have more children than did couples with a boy and a girl. Thus, since more families are ready to stop having children if they have been fortunate enough to achieve the preferred sex composition of their family, namely a boy and a girl, then it would follow that the distribution of two-child families remaining would be similar to our finding, i.e., a higher proportion of two-child, cross-sex families. Our samples were not large enough for us to test whether or not there is a differential frequency of various family constellations in families with more than two children. Westoff, however, reports that parents with three children of the same sex reported a desire for more children than did parents with cross-sex, three-child families.

Finally, it should be noted that socioeconomic differences in the families of subjects are not a primary issue when the analysis is *intra*-family. However, when *inter*-family comparisons are made, then the socioeconomic status of the families must be taken into account. This factor has been overlooked in the majority of research reporting the relationship of birth order and college attendance. In the single study which has come to our attention in which socioeconomic status was controlled (Bayer, 1966), no differences were found between first-born and last-borns with respect to attendance in college one year following graduation from high school. However, in this study there was no analysis of the sex distribution of the subjects in terms of family constellation.

To summarize, in any study of the relation of ordinal position and educational achievement, it seems of vital importance to ascertain the factor of sex of sibling in order to: (1) assess the proportion of like and opposite sex composition of the family sibling structure and (2) assess the differential contribution of same sex and opposite sex families to differences in achievement between ordinal positions. The issue of the relative proportions (or "base rates") of same and opposite sex family composition has not been dealt with in

the literature; the possibility of a differential contribution of ordinal position to cognitive factors and sex of subject and sibling to affective factors has been advanced by Sutton-Smith, et al. (1964); Koch (1954, 1955) has utilized a combination of both ordinal position and sex of sibling in her studies of personality, mental abilities, and cognitive style. In his admonition about confused definitions of sibling position, Warren (1966) restricts his comments to the categorical mixing of onlies and first-borns; this warning could well be extended to other conceptualizations which leave the reader confused as to such critical data as size of family, sex of immediately adjacent siblings, and the actual ordinal position (use of the ambiguous term, "later born," e.g.). The discussion and presentation of sibling influences on homosexuality in Bieber, et al. (1962:118–139) illustrate the omission of these critical categories; so does Bayer's (1966) investigation of the relation of birth order, socioeconomic status, and college attendance.

Our findings have led us to such criticisms of both taxonomy and conceptualization of that aspect of family structure termed sibling position. The inclusion of *all* siblings (our detailed presentation was limited to two-child families) provides the possibility for evaluation of the relevance of both ordinal position and sex of sibling to years of schooling. The inclusion of all siblings also by-passes the possibility (in the more conventional studies) that subjects from a given ordinal position came from higher socioeconomic backgrounds, so that their greater degree of schooling might be a function of family background rather than ordinal position. This is not to deny the possible influence of socioeconomic position; Warren (1966) suggests that a greater differential in schooling between siblings may occur more frequently when the family or origin is of lower socioeconomic background, while Cobb and French's (1966) findings indicate that the ratio of first-borns to later-borns in medical school is greater where there is status incongruence with the father. Knowledge of the terminal education of all siblings also permits further investigation of such specific research questions as to what the differences are between (1) families of an FM sibling structure where the girl attains more schooling than her brother and (2) families with a comparable sibling structure and socioeconomic status where the boy attains more schooling than his sister.

Summary

A more detailed analysis of factors contributing to the overrepresentation of first-borns in college populations is afforded by comparisons of years of schooling *between* siblings. Analysis of educational differences within two-child families reveals that: (1) first-borns attain significantly more education than their last born siblings *only* when the first- and last-born are of the opposite sex and (2) there is an overrepresentation of the cross-sex two-child family in three independent samples. The widely reported overrepresentation of first-borns in college is thus a function of not only ordinal position, but also sex of the subject and sex of the sibling.

References

Aberle, David F. and Kaspar D. Naegle, 1952, "Middle-class fathers' occupational role and attitudes toward children," *American Journal of Orthopsychiatry,* **22**, (April):366–378.

Altus, William D., 1965, "Birth order and academic primogeniture," *Journal of Personality and Social Psychology,* **2**, (December):872–876. 1966, "Birth order and its sequelae," *Science* (January):44–48.

Bayer, Alan A., 1966, "Birth order and college attendance," *Journal of Marriage and Family Living,* **28** (November):480–484.

Bayley, Nancy, 1966, "Methodological problems in longitudinal research," Read at the Sixth International Congress of Child Psychiatry, Symposium on Problems of Research and Methodology, Edinburgh, Scotland, July 24–29.

Bieber, Irving, *et al.* 1962, *Homosexuality.* New York: Basic Books.

Brim, Jr., Orville G., 1956, "Family structure and sex role learning by children: a further analysis of Helen Koch's data," Sociometry **21** (March):1–16.

Cobb, Sidney and John R. French, Jr., 1966, "Birth order among medical students," *Journal of American Medical Association,* **195** (January):312–313.

Jones, Harold E., 1939, "Procedures of the adolescent growth study," *Journal of Consulting Psychology,* **3** (November-December):177–180.

Koch, Helen L., 1954, "The relation of 'primary mental abilities' in five- and six-year-olds to the sex of the child and characteristics of his sibling," *Child Development,* **25** (September):209–223.

Koch, Helen L. 1955, "Some personality correlates of sex, sibling position, and sex of sibling among five and six-year-old children," *Genetic Psychology Monographs,* **52** (August):3–50.

Macfarlane, Jean W., 1938, "Studies in child guidance: I. methodology of data collection and organization," Monograph of the Society for Research in Child Development, **3**, Number 6.

Schacter, Stanley, 1963, "Birth order, eminence and higher education," *American Sociological Review,* **28** (October):757–768.

Siegel, Sidney, 1956, *Nonparametric Statistics.* New York: McGraw-Hill.

Sutton-Smith, Brian, Roberts, John M. and B. G. Rosenberg, 1964, "Sibling associations and role involvement," *Merrill-Palmer Quarterly of Behavior and Development,* **10** (January):25–38.

Westoff, Charles F., Potter, Jr., Robert G. and Phillip G. Sagi, 1963, *The Third Child.* Princeton University Press.

Joe Molnar

PART 3

Mental Health of Learners

3.1 Interpersonal Relations and Mental Health in the Classroom

RICHARD A. SCHMUCK, MARGARET B. LUSZKI, AND DAVID C. EPPERSON

Reprinted from *Mental Hygiene,* 1963, **47**, 289–299, with permission of the authors and the National Association for Mental Health, Inc. Richard A. Schmuck is professor of educational psychology at the University of Oregon and a research associate at the Center for Advanced Study of Educational Administration; Margaret B. Luszki is on the faculty of the Department of Psychiatric Medicine at the University of South Carolina; and David C. Epperson is professor of education and urban affairs at Northwestern University.

The purpose of this research was to clarify the nature of the relationship between classroom interpersonal relations and the mental health and academic learning of the pupils. The findings indicate that one of the important ingredients in achievement of academic tasks is the classroom atmosphere. The person who has the most influence and control of classroom atmosphere and interpersonal relations is the teacher.

The increasing complexity of conditions under which people in our society live and work has emphasized the importance of effective interpersonal relations. Modern life, particularly in our urban centers, places a premium on the ability to get along with others. In addition, individuals must be prepared to deal with tensions and conflicts—not merely to avoid them but to handle them constructively and creatively—if we are to solve some of the problems of contemporary society.

As a result of changes in our society during the last few years, schools have a heavier responsibility than ever before to help pupils develop behavior patterns which equip them to fill useful roles in society and contribute maximally to group productivity. This means that concurrent with the teaching of subject matter, schools must be concerned with the development of interpersonal relationship skills and positive mental health.

The question has been raised repeatedly, "What is optimal mental health?" Many definitions have been posed and many systematic lists prepared which attempt to delineate criteria or state the characteristics by which mentally healthy people can be identified. One such comprehensive attempt was made by Jahoda (3), who reviewed selected literature on mental health and derived an extensive list of factors which are viewed as indicators of mental health. These were classified into six general areas: (1) Positive attitudes toward the self; (2) Optimal growth, development, and self-actualization; (3) Psychic integration; (4) Personal autonomy; (5) Realistic perceptions of the environment; (6) Adequate environmental mastery.

The research being conducted currently at the Institute for Social Research is concerned with several of these mental health factors. This paper deals with the following:

1. *The pupil's attitudes toward himself.* The child who feels he is liked, valued, and accepted by his classmates, who describes himself in favorable terms, and who feels that he is a part of the classroom group, may be thought to be in a positive state of mental health.
2. *The pupil's perception of reality.* The pupil whose perceptions of the classroom are relatively free from distortion has better mental health than the pupil who distorts reality frequently.
3. *The pupil's mastery of his environment.* The child's adequacy in meeting the school's formal learning requirements represents one type of mastery of the environment, while his adequacy in establishing positive relationships with other pupils is another type. Both generally result in a satisfying state of affairs for the pupil and are considered to be indicators of positive mental health.
4. *The pupil's actualization of his potential.* A child with academic abilities he does not use is presumed to have poor mental health. In many such cases, energy is being drained off by excessive anxiety, worry, and hostile feelings, so that the pupil is not free to utilize his ability in performing classroom tasks.

These four elements of mental health refer to the pupil's relationship to and his feelings about people and tasks in the classroom environment. *Mental health here refers to the adequacy of the pupil's relationship to his learning environment and the positiveness of his feelings about himself.* With this orientation to mental health it is evident that the teacher is indeed in a position to influence a pupil's mental health, either positively or negatively. If the teacher possesses the skills necessary to establish a wholesome interpersonal atmosphere and if he can organize learning tasks in a manner which will enhance a pupil's self-esteem, he is contributing, simultaneously, to the improvement of a pupil's mental health.

It seems clear that mental health is influenced to a large extent by the way important people in an individual's environment respond to him. This appears to be true for very young infants, children and adolescents, as well as adults of all ages. An individual's concept of himself is built up through the accumulated reflected appraisals of those with whom he comes in contact.

Every person, in other words, makes use of the reactions of other people in formulating his opinion of himself. He relies on others, to a great degree, for the gratifications and rewards which make him feel worthwhile and esteemed, or for the punishments and disapprovals which make him feel inadequate and worthless. It is primarily other people—in person or in the images one holds of them—who are able to make an individual feel secure and happy or lost and unhappy.

Psychological research indicates that when a person feels anxious or fearful in the presence of another, he has difficulty in accurately perceiving the world. His perceptions may become so distorted that he is unable to behave appropriately. Those who have studied these processes have found that the greater the threat a person feels in the presence of another, the more pronounced the restricting and distorting effect is on his thoughts and perceptions of his surroundings.

An experiment performed by Coombs and Taylor (1) illustrates this phenomenon. Mild degrees of personal threat were introduced to students by belligerent examiners while the students were performing a task requiring intellectual functioning. The researchers predicted that this personal threat would result in an increase of time required to complete the task, as well as an increase in errors in performance.

The 50 participants in this experiment were given the task of translating sentences into a simple code. With only a single exception, the students required longer time periods to complete the coding procedure when they were working under threatening conditions, and they also made a greater number of errors of translation than in a comparable, nonthreatening condition. It is not difficult to predict what might happen to a student who again and again is presented with situations that are threatening him. The extent to which he actualizes his academic potential is likely to be considerably reduced.

Using some of these general notions, social scientists at the Institute for So-

cial Research have systematically collected information concerning the social atmospheres of classroom settings and the mental health conditions of individual pupils.

Van Egmond (6), for instance, in a study of 640 elementary school children, sought to link ideas concerning a pupil's relationships with other pupils and the extent to which that pupil actualizes his academic potential. He constructed an index of the level of actualization from the relative performance on intelligence tests and achievement batteries. He found that boys who achieve some self-esteem and recognition by being able to influence other boys come closer to actualizing their potential than boys who are not able to gain esteem and recognition through influencing other boys. For these latter boys, conditions of threat undoubtedly prevail when they are in the presence of more powerful, influential boys.

For girls, Van Egmond found that those who were liked by their classmates actualized their potential rather fully, while those who were disliked or liked by very few tended to actualize their potential less fully. For girls, the threatening circumstances of being surrounded by other girls who exhibit dislike for them appears to be disruptive enough to affect their classroom performance significantly.

In another study, Lippitt and Gold (4) found that pupils who are highly liked by others express more liking in their ratings of peers than those who are not liked by their peers. Interestingly, through the course of the school year, this difference in the feelings expressed by liked and disliked children becomes even greater, and the disliked children show more negative feelings toward their fellow classmates at the end of the year than at the beginning, while the highly liked show more positive feelings.

Some systematic observation of pupils' interaction in the classroom also supports these findings. The researchers found that those behavior patterns which indicate aggressive-assertive or passive-hostile activity are more frequently characteristic of the disliked children. These children tend to behave in ways that are likely to disrupt classroom functioning to an increasing degree as they realize how their peers feel toward them. Thus, as pupils are placed in threatening social atmospheres, they tend to react in irrational and nonadaptive ways. Often, by behaving aggressively or with hostility, the disliked pupil creates more dislike for himself, and this can go on and on in a circular fashion. The very nature of the situation must be altered before the individual can hope to regain or build up self-esteem.

In still another study of seven elementary classrooms, Echelberger (2) analyzed teacher ratings of children. These ratings were concerned with conditions of positive and negative mental health, as exhibited by the individual pupil. She found that the more influential and popular children impress their teachers with a significantly more favorable "mental health" picture. They show, for instance, fewer behavior problems, greater social adjustment, and more stable emotional patterns.

The purpose of the current research is to study various aspects of the social atmosphere of the classroom and to

relate these group relevant aspects to individual mental health. The data presented in this paper are from four upper elementary classrooms in and around Ann Arbor, Mich. These classrooms are part of a broader study of the classroom atmospheres, covering grades four through twelve, in different types of communities in southeastern Michigan.

The particular classrooms used in this analysis were selected because they represented the highest grade in which the pupils were working with the same teacher and classmates throughout most of the school day (sixth grade), and because they also represented two contrasting types of social atmospheres. The four teachers involved were all experienced and well-qualified and were considered to be among the better teachers in their particular school systems.

Propositions and Results

Classroom affection

Classroom groups, like all oher groups, have both formal and informal aspects. The formal aspects have to do with the way the various members work toward carrying out the official or specified goals of the group. In the classroom group, for instance, one formal feature is the way in which any child performs the role of pupil, as it is defined by the teacher and the adult community at large.

The informal aspects of a group, on the other hand, have to do with the manner in which each member relates to other members as persons. In the classroom group one informal aspect is the way affection, or pupils' liking for one another, is distributed. These informal features of a group have an important bearing on the formal aspects, or the way the stated objectives of the group are carried out. Many of them, such as the amount of liking members have for one another or their willingness to help and support each other, may be thought of as conditions for positive or negative mental health.

At least two kinds of informal patterns or group structures of classroom affection can be described. The first of these is referred to as a "centrally structured group," or one with a narrow focus of affection. In such a group, a large number of members agree in selecting a small group of individuals as the ones whom they "like the most," and they also agree in selecting a few other members as the ones whom they dislike. As a result of the narrow focus on a very few members who are most popular and another few members who are most unpopular, there are many members who are neglected and are mentioned by no one as being either liked or disliked.

A second kind of informal pattern is referred to as a "diffusely structured group," or one with a wide focus of popularity. In this kind of group there is a fairly equalized distribution of positive and negative choices. Here almost everyone is most liked or least liked by somebody. There are no distinct subgroups or cliques whose members receive a large proportion of the positive or negative preferences, and there are few neglected members.

Consideration of these two types of structures provides the basis for our first proposition.

General proposition I. *Classroom groups with a wide focus of affection (diffusely structured) lead more often to conditions of mental health than groups with a narrow focus of affection (centrally structured).*

Hypothesis 1. *The fact that a pupil is more disliked than liked by his peers is more obvious to him when the group is characterized by a centrally structured liking-disliking distribution than when it is characterized by a diffusely structured distribution.*

The structure of the group and each pupil's position in it is obtained by using a sociometric test. Each pupil is asked to designate four classmates whom he likes most and four whom he likes least. A pupil is given one "positive choice" when he is designated by another pupil as most liked, and one "negative choice" when he is designated as least liked.

A net score is obtained for each pupil by subtracting the total number of negative choices he receives from the total number of positive choices. The mean of this distribution of net scores is equal to zero in every classroom. Classroom groups distinguished by a net score distribution with a "high" variance are designated as centrally structured, while groups distinguished by a net score distribution with a "low" variance are designated as diffusely structured.

A pupil's estimated or perceived position in this sociometric structure is measured in the following manner: Each pupil is asked, "Where would you place yourself in judging how much others in the class like you?" He answers by checking the quarter of the class in which he thinks he belongs. For this analysis, a pupil is designated as occupying a high position if he checks either the first or second quarter, and a low position—that is, considering himself relatively unpopular—if he checks the third or fourth quarter. Table 1 indicates support for Hypothesis 1.

Hypothesis 2. *Pupils in diffusely structured groups evaluate themselves more highly than pupils in centrally structured groups.*

The pupil's self-evaluation or the way he feels about himself, is obtained from a sentence completion test. The test, as a whole, consists of 46 sentence stems, four of which relate to feelings about the self, as for example, "When I look in the mirror, I . . . . . , and "Sometimes I think I am . . . . . ."

TABLE 1. Accuracy in Perceiving Low Liking Status in Two Types of Liking Structures [a]

Accuracy in Perceiving Low Liking Status	Type of Liking Structure	
	Central	Diffuse
Accurate	20	8
Inaccurate	11	22

[a] Chi-square = 8.80 ($p < .005$) $df = 1$.

TABLE 2. Two Types of Liking Structures and Self-Esteem [a]

Self-Esteem	Type of Liking Structure	
	Central	Diffuse
High	18	34
Low	35	29

[a] Chi-square = 4.66 ($p < .05$) $df = 1$.

These are rated on a 7-point scale (5) and combined into a self-esteem index. Table 2 indicates support for Hypothesis 2.

General proposition II. *When a pupil thinks other pupils do not like him, his mental health is likely to be less positive than when he feels he is liked.*

Hypothesis 3. *Pupils accurate in estimating that they are not well-liked are lower actualizers of academic resources than pupils who are accurate and well-liked.*

A pupil's actual liking status in the classroom social structure is derived from the sociometric test. After a net score is obtained for each pupil by subtracting the total number of negative choices he receives from the total number of positive choices, pupils are rank-ordered according to these scores. This distribution is divided at the median to separate high or low status pupils.

Judgment of whether a child is achieving as well as might be expected is made by the teacher. To obtain this measurement, we divide each class at the median into a higher intelligence group and a lower intelligence group, based on scores from standard intelligence tests. The teacher then divides each group into a high achieving and a low achieving subgroup.

Thus, the class is divided into four ability-achievement groups: high ability-high achievement, high ability-low achievement, low ability-high achievement, and low ability-low achievement. The two high achieving groups are considered to have a relatively high level of actualization of their academic resources, and the two low achieving groups to have a relatively low level of actualization.[1] Table 1 indicates support for Hypothesis 3.

TABLE 3. Actual Liking Status (Accurately Perceived) and Actualization of Academic Resources [a]

Actualization	Actual Liking Status (Accurately Perceived)	
	High	Low
High	30	5
Low	18	23

[a] Chi-square = 14.19 ($p < .001$) $df = 1$.

[1] This ability-achievement measure, as indicated by the way each class is divided, is the pupil's *relative* position in his class, and not his ability or achievement in relation to all pupils of a particular age or grade.

Hypothesis 4. *Pupils who perceive themselves as not being well-liked are lower actualizers than pupils who perceive themselves as being well-liked.* Table 4 indicates support for Hypothesis 4.

TABLE 4. Perception of Liking Status and Actualization of Academic Resources [a]

	Perception of How Well-Liked	
Actualization	High	Low
High	50	12
Low	31	33

[a] Chi-square = 14.23 ($p < .001$) $df = 1$.

Hypothesis 5. *Pupils who perceive themselves as not being well-liked show lower self-esteem than pupils who perceive themselves as being well-liked.* Table 5 indicates support for Hypothesis 5.

TABLE 5. Perception of Liking Status and Self-Esteem [a]

	Perception of How Well-Liked	
Self-Esteem	High	Low
High	50	16
Low	31	29

[a] Chi-square = 7.94 ($p < .005$) $df = 1$.

A pupil's conception of himself in relation to others

When a pupil perceives a lack of congruence between his own attitudes and the attitudes of others in the classroom, the environment cannot be considered to provide many opportunities for social rewards. If, for example, a pupil wants to participate actively in classroom activities but believes others in the class would not like this, he is unable to anticipate much need satisfaction. Under these conditions he is cut off from an important source of gratification. In this sense he can be considered isolated, unsupported, or alienated from potential reinforcing agents.

Since a pupil's conception of himself as a classroom member is formulated, at least in part, by the manner in which he is responded to by his teacher and his peers, one would expect isolation or alienation from them to have an impact on his conception of himself as a learner and as a person.

If a pupil does not think others in the class value him, his image of himself as a participant in classroom learning situations is likely to be negative. Since individuals tend to behave in a manner consistent with their self images, it seems probable that those individuals who see themselves as not being valued as classroom members would be less likely to actualize their potentials as learners than those who consider themselves to be valued.

General proposition III. *Pupils whose attitudes are discrepant from those they attribute to others in the classroom also see others as not valuing them. Such pupils experience conflict, which results in reduced personal effectiveness.*

Hypothesis 6. *When a pupil thinks the teacher and the other pupils do not value him, his school adjustment will be low.*

To determine how he thinks others value him, each pupil is asked to rate himself the way he thinks his teacher sees him and the way he thinks his classmates see him. These ratings are on a nine-point scale ranging from "many things others like about him" to "many things others do not like." The ratings of the teacher and the classmates are combined to form a single index.

The school adjustment items used to test this hypothesis consist of five stems from the sentence completion test such as "Studying is . . . . . ." and "This school . . . . . ." Each stem is rated on a 7-point scale (5) and a mean rating computed for each pupil. Table 6 indicates support for Hypothesis 6.

TABLE 6. School Adjustment and How He Thinks Others in the Class Value Him [a]

School Adjustment	How He Thinks Others Value Him	
	High	Low
High	29	20
Low	21	42

[a] Chi-square = 7.45 ($p < .01$) $df = 1$.

Hypothesis 7. *When a pupil thinks the teacher and the others in the class do not value him, his actualization of academic potential is reduced.* Table 7 indicates support for Hypothesis 7.

Hypothesis 8. *A lack of congruence between how the pupil feels about classroom relevant behaviors and how he thinks the teacher feels about these same behaviors is accompanied by a low level of actualization.*

TABLE 7. Actualization and How He Thinks Others in the Class Value Him [a]

Actualization	How He Thinks Others Value Him	
	High	Low
High	32	24
Low	18	38

[a] Chi-square = 7.08 ($p < .01$) $df = 1$.

In order to measure congruence between the pupil's attitudes and those which he attributes to the teacher, each pupil is given a series of statements dealing with classroom standards, as for example, "It is good to take part as much as possible in classroom work," and "If you work very hard others in this class will not like it." He is asked to indicate how he thinks others feel about each of these, putting his answer on a 5-point scale ranging from "almost everyone in the class thinks this," to "only a few in the class think this."

He is also asked to indicate how he, personally, feels, how he thinks the teacher feels, and how his best friends feel. Discrepancy scores are then obtained between his own feelings and the feelings he attributes to his classmates, best friends, and teacher. The discrepancy between his own attitudes and those he attributes to the teacher is the measure of congruence used to test this hypothesis. Table 8 indicates support for Hypothesis 8.

Hypothesis 9. *A lack of congruence between how the pupil feels about classroom relevant behaviors and how he thinks the teacher feels is accompanied by a high desire for change in the teacher.*

TABLE 8. Actualization and Congruence with the Teacher [a]

Actualization	Congruence with Teacher: High	Congruence with Teacher: Low
High	38	23
Low	25	37

[a] Chi-square = 5.94 ($p < .02$) $df = 1$.

Information regarding desire for change is obtained from the same instrument used to measure classroom standards (described under Hypothesis 8). On it, pupils are asked to indicate how they would like the class and the teacher to change in the opinions they hold. In another instrument, on which they rate the characteristics of their teacher, they are asked how they would like their teacher to change. The measure used to test this hypothesis was a combination of these two indices: desire for change in classroom standards, and desire for change in teacher characteristics. Table 9 indicates support for Hypothesis 9.

Hypothesis 10. *A lack of congruence between a pupil's own attitudes and those he attributes to the teacher is accompanied by a low attraction to the class.*

TABLE 9. Desire for Change and Congruence with the Teacher [a]

Desire for change	Congruence with Teacher: High	Congruence with Teacher: Low
High	24	42
Low	39	18

[a] Chi-square = 12.58 ($p < .001$) $df = 1$.

In order to establish the pupil's attraction to the class, each was first asked how much time he would like to spend in class. Answers were recorded on a scale ranging from "a lot more" to "a lot less." He was then asked to rank order the different parts of his day, including life in this class, to indicate their relative importance to him, the amount of happiness each brings, and the amount of learning derived from each. Finally, he was asked to check a 6-point scale which ranged from "This class with your regular teacher has mostly good things," to "This class with your regular teacher has mostly bad things." These three sources of data were combined to form an index of attraction to the class. Table 10 provides support for Hypothesis 10.

TABLE 10. Congruence with the Teacher and Attraction to the Class [a]

Congruence with teacher	Attraction to Class: High	Attraction to Class: Low
High	37	21
Low	22	32

[a] Chi-square = 5.96 ($p < .02$) $df = 1$.

Hypothesis 11. *A low attraction to the class will be accompanied by a low level of actualization.* Table 11 indicates support for Hypothesis 11.

Implications for Teachers

These findings emphasize the relevance of positive affect or liking among members of a classroom group for individual

TABLE 11. Actualization and Attraction to Class [a]

Actualization	Attraction to Class: High	Attraction to Class: Low
High	35	21
Low	24	32

[a] Chi-square = 4.37 ($p < .05$) $df = 1$.

mental health and for effective learning. When the pattern of affection is diffuse, so that almost every member is "most liked" by some other member, pupils tend to have more positive feelings toward themselves, perceive the school situation more favorably, and actualize their potentials more fully.

If, as these findings indicate, successful human relations are important to achievement of academic tasks, the teacher should do everything possible to enhance a pupil's ability to obtain emotional support from his peers, not only because good human relations is a value in itself, but also because it contributes importantly to the school's academic goals.

What do findings of this type suggest for teacher behavior? First is the job of obtaining accurate information concerning the distribution of liking choices in the classroom. One way of accomplishing this is through careful and systematic observation.

The teacher might ask himself: Are there some pupils who tend to be left out of most classroom activities? Are there some who are always first chosen, and others who are always last chosen?

He should raise questions such as these in attempting to view his class as objectively as possible, looking at it from time to time as if he were an outside observer. He might also examine his own behavior and attitudes: How do I distribute rewards and punishments in the classroom? Are there certain children to whom I give primarily negative criticism and rebuke, and others who receive a large portion of the praise? In addition to careful observation and self-questioning, the teacher might employ such sociometric techniques as those used in the research reported here in order to find out how interpersonal affect is distributed in his group.

Once the classroom distribution of affect is clear to the teacher, and if some change is desirable, he may use various techniques to modify this distribution and develop more positive feelings among the pupils. A co-operative study group, for instance, in which low and high status pupils work together for the achievement of some common goal, is quite often effective in changing inaccurate perceptions and stereotypes about low status children.

Another possibility is to work low status children gradually into roles which are viewed as having considerable status. Some teachers have accomplished this effectively by developing a so-called "steering committee" in which all pupils serve at some time during the school year and through which each achieves some status and self-esteem by being active contributors to classroom decisions.[2]

[2] A report of the case of a "steering committee" in two elementary classrooms, together with other classroom innovations, is contained in *Inventory of Teaching Innovations, Directed toward Improving Classroom Learning Atmospheres,* Institute for Social Research, University of Michigan, December, 1961.

Teachers can enhance the affective tone of the classroom by including information and discussion about the nature of individual differences as a part of the subject matter. An understanding of differences among pupils relative to family background, sex, race, abilities, and interests increases the opportunity for pupils to be more accepting and tolerant of a greater number of their peers. In addition to producing greater acceptance of individual pupils, a classroom standard of acceptance of differences may emerge from such exploration.

Another approach to developing a wider focus of acceptance is through school programs directed toward a greater understanding of behavioral causation, or the "whys" of human behavior. The assumption behind this type of curriculum content is that greater insight into those factors which contribute to a pupil's actions will result in a more accepting classroom atmosphere.

The teacher's own behavior and feeling toward individual pupils can also contribute toward the acceptance of a pupil by his peers. If the teacher himself accepts each pupil as an individual—understanding his limitations and giving him the kind of support he needs to expand his assets and help overcome his shortcomings—members of the class will tend to follow a similar pattern. As a result, there will be a climate of mutual support in the classroom. But if the teacher supports primarily the high achievers and shows rejecting or disapproving behavior to those who are not so successful in classroom learning tasks, a competitive, nonsupportive climate is likely to emerge.

The teacher can give classroom relevant rewards directly by making positive comments about the pupil's performance or indirectly by organizing learning experiences in a manner which will maximize success and reduce failure. In this way, too, he creates the anticipation of future rewards. With more experiences of success, a pupil's self-esteem is enhanced, and concurrently his perception of the school environment should tend to become more positive.

All of these suggestions involve increasing the social-emotional support for all pupils in a classroom group. A second kind of implication emerging from these research findings involves pupil-teacher and pupil-pupil communication of standards and attitudes concerning classroom relevant behavior.

It has been found that when a pupil's attitudes are discrepant from those he attributes to others, he tends to feel he is not valued, his personal effectiveness is reduced, and his school adjustment is low. Often this condition arises from poor communication—that is, when opportunities are not available for an active and public exploration of the interpersonal expectations of both pupils and teacher.

A pupil's own attitudes may actually be quite similar to those of his classmates, but because of inadequate communication he believes they are different. The teacher might be able to correct these erroneous beliefs, or misconceptions, by improving communication in the classroom through discussions of how pupils and he himself feel about learning tasks. By way of such discussions, classroom standards can be clarified and, if necessary, reformulated so

that they are made clear and are generally accepted.

Summary

This study attempts to clarify the nature of the relationship between classroom interpersonal relations on the one hand and mental health and academic learning on the other. For the purpose of this exploration, classroom mental health refers to the adequacy of the pupil's relationship to his learning environment and the positiveness of his feelings about himself. Some dimensions of this definition were suggested, and some ways of measuring mental health were described.

Three general propositions were put forward dealing with: (1) The distribution of liking and disliking choices in the classroom; (2) The pupil's perception of how well he is liked; and (3) The pupil's conception of himself in relation to others. The findings suggest that there is an integral relationship between classroom interpersonal relations and pupil mental health. Several implications for improving the classroom learning environment were made on the basis of these and similar findings.

References

1. Coombs, A. W. and C. Taylor, "The Effect of the Perception of Mild Degrees of Threat on Performance," *Journal of Abnormal and Social Psychology,* 1952, **47**, 420–24.
2. Echelberger, E., *Relationships between Personality Traits and Peer Status* (Ann Arbor: University of Michigan, 1959). Unpublished doctoral dissertation.
3. Jahoda, Marie, *Current Concepts of Positive Mental Health* (New York: Basic Books, Inc., 1958).
4. Lippitt, R. and M. Gold, "Classroom Social Structure as a Mental Health Problem," *Journal of Social Issues,* **15** (No. 1, 1959), 40–49.
5. Malpass, L. F. and F. B. Tyler, *Validation of the Incomplete Sentences Test of School Adjustment* (Carbondale, Ill.: Graduate School, Southern Illinois University, 1961). Unpublished research report.
6. Van Egmond, E. E. *Social Interrelationship Skills and Effective Utilization of Intelligence in the Classroom* (Ann Arbor: University of Michigan, 1960). Unpublished doctoral dissertation.

3.2 *Social Schema of Normal and Disturbed School Children*

RHODA LEE FISHER

Reprinted from the *Journal of Educational Psychology,* 1967, **58,** 88–92, with permission of the author and the American Psychological Association, Inc. Rhoda Lee Fisher is a clinical psychologist at the Child Guidance Clinic, Syracuse, N.Y. and also serves as a psychological consultant to a number of agencies in that area.

Fisher's study shows that children with serious school problems feel themselves to be relatively distant or estranged from others, and, furthermore, that those children with hostile mothers conceived of social relations as more distant than children whose mothers were less hostile. This study is an example of the way in which parental attitudes become reflected in the behavior of children.

The present studies concerned themselves with a group of children whose deportment was so disruptive that teachers could not exercise sufficient control over them to permit them to remain within the normal classroom. The inability to behave in an acceptable socially approved fashion seems to be a critical problem for them. A measure of social distance has been developed by Kuethe (1962) which schematically represents the individual's concept of social interaction. With this technique he found that

individuals who employ a specific social schema in organizing behavior in one situation will employ the same schema in quite different situations where the one common denominator is social stimuli of the same content even though the physical form of stimuli may be completely changed [*Kuethe, 1964, p. 24*].

Within the context of the application of the social schema technique, the following objectives were set:

1. To compare the social schema of normal children with those characterizing children with school behavior problems. It was anticipated that those showing disruptive behavior in the classroom would portray interaction schema differently than would normally behaving children. More specifically it was expected that the disturbed children would experience themselves as having fewer meaningful ties to others—as more alienated and without meaningful roles. That is, they feel separated and distant from others.
2. To relate distances set by children between human forms and the degree of hostility characteristics of their mothers. It was hypothesized that the more angry and aggressive the mother the more distantly would her child structure social relations. That is, one would expect that the angry mother would be perceived as repelling and pushing others away. She would exemplify a style of social relationship based on distance and unfriendliness.
3. To compare a group method of administration of social distance schema with an individual method of administration.

Study I

Method

Subjects. One group of subjects (*Ss*) consisted of 32 white male elementary school children in special classes designed for those unable to adjust to the regular school demands. These children were too difficult for the regular classroom teacher to control. The mean age of the group was 9.5 (range 6–13). A second group of *Ss* was comprised of 45 white male elementary school children from the regular fifth- and sixth-grade classrooms; and a third group included 44 white female elementary school children from the regular classroom. The mean age for both normal groups was 11.2 (range 10.0–13.0).

Procedure

An adaptation of a technique developed by Kuethe (1962) for the measurement

of social schemata was used. Instead of Kuethe's usual felt figures with directions to place them on a large felt background mounted on a wall, the instructions were to glue the figures on an open field (in a booklet of paper). This shift in procedure enabled the series to be administered on a group basis. Specifically, each sample of *Ss* placed eight groups of figures on a series of fields according to Kuethe's (1962) "free response instructions." The figures for the group series consisted of yellow cutouts on gum-backed paper varying in height from 2¼ inches (the baby figure) to 5½ inches (the man figure). Each *S* was provided with a booklet consisting of eight pages of 6-inch by 22-inch paper. An envelope containing figures to be placed on each page was distributed to each *S*. Not until everyone had completed his arrangement of the figures on a given page were instructions given to turn to the next page; and the next envelope was distributed. A description of the figures contained in each envelope follows:

Envelope 1: two boys.
Envelope 2: a girl, a woman.
Envelope 3: a woman, a man.
Envelope 4: two boys, one girl.
Envelope 5: a man, a girl.
Envelope 6: a man, a boy.
Envelope 7: a boy, a girl.
Envelope 8: a boy, a girl, a baby.

Scoring consisted of measuring the distance between each two figures. The score derived was the average distance for the entire eight situations.

Results and discussion

The average distance between the figures arranged in schemas was significantly different for the disturbed and the normal samples (see Table 1). Normal boys arranged the human figures more closely together than did the disturbed boys. Mean total average distance for the normal boys was 1.6 inches; mean total average distance for the group of disturbed boys was 2.3 inches ($t = 2.9$, $p < .01$).

A comparison between the normal boys and the normal girls failed to reveal a real difference. The total average distance of the figures as grouped by the girls was 1.6 inches.

TABLE 1. Means and Standard Deviations of the Normal and Disturbed Children on the Group Administered Form of the Social Schemata Technique

Subjects	Total Average Distance		Average Distance between Child Figures	
	M	*SD*	*M*	*SD*
Disturbed boys	2.3	1.2	2.3	.7
Normal boys	1.6	1.0	1.7	1.1
Normal girls	1.6	.9	1.9	1.4

Since no sex differences were apparent between the normal boys and the normal girls it seemed logical also to compare the normal girls with the disturbed boys. When this was done, a significant difference was once again found ($t = 2.3$, $p < .05$).

Another analysis in which age was controlled indicated that age exerted no influence on the distance between figures as represented in the social schemata. Children below the age of 10.0 were eliminated from the disturbed group. Children above the age of 11.0 were eliminated from the normal group of boys. The mean age of both groups for this analysis was 10.4 years. Mean total distance between figure placement for the normal boys ($N = 12$) was 1.4 inches. The closeness of these values to those for the total samples indicated that age did not influence the perception of social distance in the present experimental context.

As predicted, normal children placed figures with human characteristics closer together than did children with serious school problems. In terms of Kuethe's work this suggests that the normal children feel closer and more related to others than do the disturbed children.

In a paper recently published Weinstein (1965) also reports that the Kuethe schema placements discriminate disturbed and normal boys. The figures and felt size used by Weinstein were not directly comparable to those employed in the present study. In her study a group of 20 disturbed boys admitted to a special residential school for emotionally disturbed children was evaluated. A group of 20 normal control children from a nearby public school matched in age and IQ was used as a comparison group. Results indicated that disturbed boys placed figures that represented mother-child relationships at a relatively greater distance from each other than did normal boys. A second finding indicated that disturbed boys, as compared to normal boys, visualized human relationships as involving more intervening distance than is true of relationships between abstract geometric forms. Overall the Weinstein data and those from the present project, while not specifically supportive of each other, certainly indicate in common a trend for disturbed boys to conceptualize human relationships as distant.

Study II

Method

For this study only those boys in the special disturbed classes were used. It was the purpose of this project to investigate relationships which might exist between the hostility attributes characterizing the mother and the closeness-distance positions assigned to human figures on the Kuethe schemas by the children. Since aggression is such a prominent dimension in the behavior of the children it seems an important dimension to investigate in the mothers as well. To this end, five subscales of the Buss-Durkee Hostility Scale (Buss, 1961, pp. 171–172) were administered to each mother. These five scales measured Assaultiveness, Irritability, Negativism, Verbal aggressiveness, and Guilt. The *S* was asked to respond with a true or false answer to

48 different statements regarding the expression of anger. Examples of the Assaultive aggressive subscale are "Once in a while I cannot control my urge to harm others," "People who continually pester you are asking for a punch in the nose," "I have known people who pushed me so far that we came to blows." Examples of the Irritability scale are "If someone doesn't treat me right, I don't let it annoy me," "Sometimes people bother me just by being around," "I can't help being a little rude to people I don't like." The Negativism scale includes statements such as "Unless somebody asks me in a nice way, I don't do what they want," "When I disapprove of my friends' behavior, I let them know it," "I can't help getting into arguments when people disagree with me" are examples of the Verbal aggression subscale. The Guilt subscale includes statements such as "I sometimes have bad thoughts which make me feel ashamed of myself," "I am concerned about being forgiven for my sins."

An individual form of the social schemata was used in this experiment with the disturbed group of boys. The individually administered form consisted of seven groups of figures cut from yellow felt. The figures were derived from Kuethe's (1962, pp. 32–33) original patterns and *Ss* were instructed to arrange each group of figures on a black felt field 36 inches by 36 inches that was stretched on a wall of the experimental room. This series consisted of the following groups:

1. Two children and a square.
2. Two children and a woman.
3. Two children and a man.
4. A man, a woman, a square, and a large rectangle.
5. A man, a woman, a child.
6. Six children.
7. A child, a book, a circle.

Results and discussion

Four scores were obtained from the individually administered technique: the total average distance between all of the figures, the average distance between all of the adult figures, the average distance between all of the child figures, the average distance between adult and child figures.

Two measures from the Buss-Durkee Hostility Scale were positively related to the distance placement between figures in the social schemata, as Table 2 shows. The Buss-Durkee Assaultiveness scores of the mothers were positively correlated with the children's indicated distances between adult Kuethe figures ($r = .57$, $p < .01$, $N = 25$). The mothers' Irritability scores were positively correlated with the distances between adult Kuethe figures ($r = .53$, $p < .01$, $N = 25$). The total average social schemata distances were also positively correlated with the mothers' Irritability score ($r = .53$, $p < .01$, $N = 26$). Mothers' Irritability scores were also positively correlated with the distances between the child figures ($r = .37$, $p = .05$, $N = 26$).

Generally then, children who place human figures at a relatively large distance from each other have mothers who are depicted as angry and hostile.

TABLE 2. Correction of Mothers' Buss-Durkee Hostility Scale Scores with Various Measures of Kuethe's Social Schemata

Buss-Durkee	Individually Tested			Group Tested
	Average Adult Distance ($M = 2.9$)	Average Child Distance ($M = 3.1$)	Total Average Distance ($M = 3.1$)	Total Average Distance ($M = 2.3$)
Direct aggression ($M = 3.6$)	.57**			
Irritability ($M = 5.2$)	.53**		.55**	
Negativism ($M = 1.8$)		.37*		—.68**

*$p = .05$.
**$p < .01$.

Study III

Method

A comparison of the two forms (individual and group) of administration of the social schemata technique was the purpose of the third study. The children in the special classes for the disturbed were administered both versions of the Kuethe procedure. Though the situations for each of the series are not completely parallel, they seemed sufficiently similar to make such an analysis worthwhile.

Results and discussion

Distance between adult figures for the individually administered schemata was significantly related to the equivalent distance between figures for the group-administered schemata ($r = .43$, $p = .02$, $N = 30$) (see Table 3).

Distance between adult figures for the individually-administered schemata was significantly related to distance between child figures for the group-administered schema ($r = .37$, $p = .05$, $N = 30$) (see Table 3).

There is evidence indicating that a similar dimension is tapped by both versions of the Kuethe technique.

Conclusions

Children with problems in adjustment to the classroom framework place human figures at a significantly greater distance apart than do children who are able to adjust successfully to the classroom. In terms of Kuethe's findings (1962), this indicates that the disturbed children feel relatively distant or estranged from others. Whether this sense of distance is produced by the negative consequences of their behavior or actually plays a previous role in contributing to their poor adjustment remains to be determined. There is the possibility that the degree of distance felt by a child between himself and others arises from socialization experiences. This is

TABLE 3. Correlations between the Group Administered and Individually Administered Forms of the Social Schemata Technique

Individually Administered Form	Group Administered Form: Total Average Distance ($M = 2.3$)	Group Administered Form: Average Distance between Child Figures ($M = 2.3$)
Average distance between adult figures ($M = 2.9$)	.43**	.37*
Average distance between adult and child figures ($M = 2.9$)	.09	.08
Average distance between child figures ($M = 3.1$)	.05	.03
Total average distance ($M = 3.1$)	.18	.12

*$p = .05$.
**$p = .02$.

suggested by the fact that a significant positive correlation was found between Kuethe scores and measures of mother hostility in the disturbed group. Structuring one's relations in a "distant" fashion may be a consequence of dealing with a mother who angrily pushes him away or perhaps of needing to remain far enough away from mother to avoid her aggression.

The Kuethe social schema technique significantly differentiated between emotionally disturbed and normal boys of elementary school age. The difference was such that the disturbed boys placed greater distance between figures in the social schemas. The degree of distance the disturbed boys place between the figures was positively correlated with the amount of hostility shown by their mothers. Mothers' hostility was measured with the Buss-Durkee Hostility Scale. It would appear that children with hostile mothers conceive of social relations as more distant than children whose mothers are less hostile.

References

Buss, A. H. *The Psychology of Aggression.* New York: Wiley, 1961.

Kuethe, J. Social schemas. *Journal of Abnormal and Social Psychology,* 1962, **64**, 31–38.

Kuethe, J. Pervasive influence of social schemata. *Journal of Abnormal and Social Psychology,* 1964, **68**, 248–254.

Weinstein, L. Social schemata of emotionally disturbed boys. *Journal of Abnormal Psychology,* 1965, **70**, 457–461.

3.3 *The Disturbed Child in the Classroom*

KATHARINE F. TIFT

Reprinted from the *NEA Journal,* March 1968, **57** (3), 12–14, with permission of the author and the National Education Association. Katharine F. Tift teaches English and Civics at Kensington Junior High School, Montgomery County, Maryland, and was formerly an educational therapist with the Hillcrest Children's Center, Washington, D.C.

Tift's article describes ways in which a teacher can keep a disturbed child from destroying the learning climate in the room and interfering with classroom goals, while at the same time helping him to understand his unacceptable behavior.

If you are assigned a student whose daily actions foul up the learning climate in your classroom, you must have feelings of frustration, and even of fury, at times. The purpose of this article is, first, to reassure you that you are *not* inadequate because you wish the child would disappear forever and, second, to suggest specific ways for dealing with problems which arise because he does not disappear.

One note of clarification: If the comments which follow seem overly simple, it is because emotional illness is overly complex. Little attempt will be made here to analyze the causes of emotional disorder; rather, the goal is to provide practical suggestions for dealing with those surface behaviors which disrupt your classroom.

Try to visualize a "typical" disturbed child. DC is a ten-year-old whose social behavior much of the time is at a three-year-old level. The minute your back is turned he runs his pencil across a neighbor's worksheet, he dips water from the fishbowl into the clay barrel, or he shoves and trips others without warning. He keeps you on edge because of his destructive actions but masters just enough subject matter and conforms just enough to your demands to keep you from telling your principal: "Either he goes or I do." (You can't ask to have him transferred; the other teachers have DC's too!)

When you first get such a student, a useful step is to check with other staff members who have come in contact with him. Does your school counselor have a case record for him? Did *other* teachers find your DC disruptive in *their* group situations?

If their answers are No, put aside this article and, instead, ask a peer whom you trust to visit your classroom for a day. Tell him to observe *your responses* to the youngster, rather than vice versa. If your friend's findings lead you to seek guidance for *yourself,* be thankful for the experience! As you acquire insights into your own feelings, everyone—especially you—will be the winner.

Once you feel confident, however, that *you* are not largely responsible for this "enemy" action in your classroom, take the next step of acquainting yourself with the general nature of DC's disease. What *is* emotional disturbance? Where does it come from? In what different ways does it manifest itself? For meaningful answers to these questions, try the following activities:

1. See the motion picture, *David and Lisa.* (Your principal might arrange a showing for his whole staff.)
2. Visit an accredited school for emotionally disturbed children. (Observe there for a day and talk with different staff members.)
3. Read at least one book which deals with the emotionally disturbed child in the classroom. (In my opinion, one of the best texts to date is *Conflict in the Classroom: The Education of Emotionally Disturbed Children,* edited by Long, Morse, and Newman. Wadsworth, 1965.)

As you follow up these suggestions, you will become increasingly aware that the behavior which angers you is a *signal* of an illness, just as red spots are a signal of measles. You will understand that a disturbed child does not

have a disease of destroying property or hurting people; these are but symptoms of his emotional disorder. And you will be reassured, once again, that *you are not responsible for his illness.*

The completion of step one (ascertaining that you'll be part of the solution, not the cause of the problem) and step two (becoming aware that emotional illness has deep environmental roots, that it is very complex, and that you alone won't "cure" it) frees you to ask the constructive "How can I help this DC and thus help all of us in the group?" instead of the defensive "Why is he doing this to *me?"*

For a third positive step, read about the following classroom situations, imagining yourself as the teacher in each of the three.

Situation number one: As you sit with a reading circle, you see the children stop reading and begin to stare across the room at DC, who is drawing pictures on the floor with chalk.

Your response to help DC goes something like this. You say clearly, for all to hear, "DC, I'm unhappy because my reading students are looking at your pictures instead of at their books. Tim [Tim is a dependable "normal"] will you help DC erase the chalk from the floor and then you and he may carry some books to the library for me."

Comments:

1. DC is relieved because (a) his unacceptable behavior was clearly defined, and (b) provisions were made enabling him to stop the behavior.

2. Tim experiences an ego boost. He has helped another human, and thus his concept of self-worth is enhanced.

3. Other class members feel secure: DC's behavior did not take their teacher from them.

4. You enhance *your* self-concept by dealing constructively with a challenging situation.

Question: But DC had unmet needs! Shouldn't he have been allowed to work them out on the floor with the chalk? The pictures weren't hurting anything!

Answer: Nonsense. DC's behavior was destroying the learning climate in your classroom. Furthermore, if DC had been permitted to continue unchecked, he would have experienced progressive feelings of:

Guilt ("Teacher is mad at me. I'm bad.")

Panic ("Help! I'm losing control!")

Anger ("Why won't someone *stop* me?"

Situation number two: DC keeps interrupting a small discussion group. You feel the children's annoyance at him, and this intensifies your own impatience with his behavior.

You let the tone of your voice, as well as your words, communicate the group indignation. "DC, you're butting in and taking other people's turns. I don't like to have you take over while I'm talking! Now, I want you to get the card box so we can help you take turns. Okay?"

You know from experience that your students all enjoy this game. The card box referred to holds about 50 blank 3 x 2 cards. Each person receives the same agreed-upon number of cards, and every time a student speaks he must put one of his cards back in the box. When someone's cards are used

up, he cannot talk until everyone else's cards are gone. (The students learn through trial-and-error how many times they are willing to permit DC to speak, and thus commit *themselves* to speaking.)

Comments:

1. DC is grateful that his unacceptable behavior was identified with honest directness. (A disturbed child cannot tolerate a honey-sweet, "Let's not do that anymore," ladled out through clenched teeth. It comes through to him as, "I'm pretending to like you, but only because I'm afraid of my real thoughts. I wish you'd drop dead.") The other students are also grateful because your acknowledgment of resentment allows them to feel comfortable with, instead of guilty about, *their* anger toward him.

2. Again you've given prompt assistance with impulse control, involving other group members. By playing a game that structures "equal rights," the pupils have a pleasurable as well as constructive role in helping their classmate.

3. You feel better and better. As you involve everyone in helping DC, it becomes "our" class instead of "my" class. A family bond begins to evolve.

Question: But DC is getting extra attention. Is that fair?

Answer: Who said life is fair? An emotionally ill person is a dependent person and requires extra attention. Don't you, as an adult, pay extra taxes to provide for patients in prisons and mental hospitals?

Situation number three: DC refuses to do a written assignment at his desk.

Withdrawal behavior presents a special challenge. Your first response might well be to hypothesize *why* he is refusing:

Too short an attention span? Call on a dependable pupil: "Jill, here is my stopwatch. Would you see how long it takes DC to complete his worksheet? DC, see if you can finish it before ten minutes are up."

He just can't do the work? "DC, I want you to do at least the first sentence now. I'll sit here and help you with it."

He simply won't do the work? "DC, I'm sorry you aren't ready to write today. Maybe tomorrow you'll be able to." (Coaxing or pleading will only reinforce the behavior you wish would go away.)

Comment:

Of course there's a chance that none of these approaches will work. If he shows total resistance, keep your cool and retreat for awhile.

Question: But doesn't that mean he wins?

Answer: Wins? Who declared war? This child is *not* attacking you; he is just protecting himself from real or imagined danger.

Question: Well, after I've retreated for awhile, should I try again to get him involved?

Answer: Of course. Does a doctor make out a single prescription, and then abandon his patient if it doesn't work?

In the above three incidents, you responded differently to different problems, but your approach each time included:

1. Stating clearly to the child what his inappropriate behavior consisted of

2. Identifying your own feelings about this behavior
3. Providing a supportive structure for a change of behavior
4. Using, whenever appropriate, the participation of other students in this supportive structure.

Incorporating these basic steps, you can deal with a variety of disruptive classroom situations. DC could be a girl instead of a boy, of course. (While statistically we have more disturbed boys than girls, how many giggly girls disrupt a lesson because they were brought up as "first a female" while they heard their brothers challenged to be "serious students" first of all?)

Or instead of being age 10, DC could be 4 or 14. (Although he may be emotionally ill, a high school student must have mastered a fair amount of impulse control to survive in that structure.) Also, DC's behavior could take the form of *repression* of hostile feelings rather than expression of them. (A silent resister can sometimes hold up group effort more effectively than the loud-mouthed extrovert.)

(One unholy situation which has not been discussed is what to do when a *number* of DC's are placed in your already overcrowded classroom. In my opinion, you'll have to give up. The only question is how you go about throwing in the sponge.)

In the foregoing discussion, our disturbed child was compartmentalized as though he were much different from his peers. In reality, no clear line can be drawn between "healthy" behavior and that which isn't so healthy. Each of our students moves up and down his own continuum of neurotic responses which he employs to master a particular environment at a particular time. But the disturbed child makes *compulsive responses* which occur *day after day* and which *interfere regularly* with classroom goals.

While there are constructive ways to cope with a DC, such as those suggested above, as long as a teacher is asked to contain sick psyches while he teaches subject matter, he will have cause for deep frustration. Happily, there is help on the horizon in addition to that which may be available to your school system from experts in special education. One form this help is taking is the study carrel, with CAI (computer assisted instruction). R. Louis Bright of the U.S. Office of Education predicts that within another decade "almost the entire academic portion of instruction will be on an individualized basis in most schools."

If this forecast becomes fact—and if subject matter is fed into machines, to be reached for by each learner as he becomes ready to assimilate it—DC's behavior will no longer stop other pupils from gaining academic instruction.

No longer will you, the teacher, face the dilemma of *"his* needs or *theirs?"* Rather, with the help of educational technology, you will have added time for those caring relationships which every human hungers for and which bring to him a sense of self-worth. You will have more time, through small-group encounters, to help each of your students experience individual growth by making important contributions to the needs of others.

3.4 *Effects on Cheating of Achievement Anxiety and Knowledge of Peer Performance*

JEV SHELTON AND JOHN P. HILL

Reprinted from the *Proceedings of the 76th Annual Convention, American Psychological Association,* 1968, **3,** 349–350, with permission of the authors and the American Psychological Association. At the time this article was written, Jev Shelton and John P. Hill were both associated with the Institute of Child Development at Minneapolis. John P. Hill is now with the Department of Human Development and Family Studies at Cornell University.

This research with high school students was designed to study the effect on cheating when students were told of the performance of another group of high school students. It was learned that such knowledge of peer performance did induce cheating, but only with students with moderate and high anxiety levels. With such students it did not matter whether they were told that the other peer group did better or worse than they. The authors hypothesized that such students always expect aversive consequences in spite of their own ability.

Three recent studies implicate achievement anxiety and knowledge of the performance of successful peers in the instigation of deviant achievement responses. Gilligan (1963) found that test anxiety was positively associated with cheating on a ray-gun game in which success is impossible without commission of proscribed responses. The test anxiety scores were interpreted as a measure of fear of failure, with cheating seen as a means of avoiding failure. Using the same task, Hill and Kochendorfer (1966) and Shelton (1966) found that the incidence of cheating was significantly greater when subjects (*S*s) were informed about the successful performance of peers than when they were not. These and associated findings have led to the general hypothesis that (*a*) the provision of information about successful peer performance in the presence of failure leads to the anticipation of aversive social consequences; and (*b*) that cheating is a response instrumental to the avoidance of those consequences. Achievement anxiety scores could then be related to cheating because they reflect a generalized expectancy of such negative consequences in the face of failure.

The present study is directed toward obtaining information about the generality of the results on which this hypothesis is based, pursuant to more direct future tests of the hypothesis. Achievement anxiety and knowledge of peer performance have not been studied concurrently. While significant main effects for both variables are predicted on the basis of the past studies, it may be that the effect on cheating of knowledge of peer performance is stronger in, or is present only for, high-anxious *S*s; or that anxiety scores are related to incidence of cheating only when information about peer performance is provided.

The studies using the ray-gun provide an external standard against which *S* can measure his own performance: a stated number of points is required for *S* to win a badge. The experimenter controls performance outcome so that no *S* wins legitimately. Would the same effects on cheating be obtained if *S*'s failure were determined solely on the basis of a standard provided by the knowledge of peer performance superior to his own? It may be, too, that the possibility of social comparison itself, rather than failure, leads to the anticipation of aversive consequences. Adding a condition, in which *S*s are exposed to information about peers who have done less well than they have, could therefore provide valuable information. If *S*s who are "successful" relative to peers are as likely to cheat as those who "fail" relative to peers, then knowledge of peer performance itself, rather than favorable and unfavorable comparisons, would be implicated in the instigation of cheating. The present study was designed to obtain information about these issues by implementing Success, Failure, and No Peer Information (Control) conditions in an academic achievement setting and measuring achievement anxiety in *S*s.

Method

The 49 boys and 62 girls in 5 sophomore and junior college preparatory classes in 2 high schools participated

in this study. Two male graduate students served as *E*s. The procedures were introduced to *S*s as tests measuring creativity. In the first session, *S*s in their regular classrooms were administered 3 tasks in the following order: (*a*) construction of as many English words as possible in 8 minutes from the letters in the word "generation"; (*b*) the Achievement Anxiety Test (AAT; Alpert & Haber, 1960); and (*c*) the first 6 items of a story-writing task. The latter task functioned as a filler in order to provide a rationale for returning the following day for a second session.

During the intersession interval, the number of words written by each *S* was recorded and the AAT was scored. Three classrooms were arbitrarily selected for the manipulation of knowledge of peer performance. The *S*s were listed in rank order of scores on the Debilitating Anxiety Scale of the AAT and, in order to equate experimental groups for anxiety level, were assigned serially from this rank-ordering to Success or Failure conditions. Implementation of these conditions required the preparation of information to be given *S* about the "performance" of a reference group. In order to hold constant the discrepancy between *S*'s performance and the alleged reference group's performance, the latter was keyed to each *S*'s actual performance. For the Success condition, average reference group performance was established at a level from 5 to 7 words below the number of words *S* wrote. Similarly, for the Failure condition, average reference group performance was established at a level from 5 to 7 words above the number of words written by *S*. For the remaining two classrooms (Control condition) no reference group information was prepared.

In the second session, *S*s completed two additional story-writing items. Then the test forms containing the words constructed from the word "generation" were returned unmarked and apparently undisturbed. In addition, a complete list of the words possible from the word "generation" was distributed together with purported average reference group performance, if any, appropriate for each *S*. The latter was embedded in a short statement which described the reference group as consisting of "good students from a good high school" and indicated that students who wrote that many words tended to do well in college. It was suggested that by scoring their own papers *S*s would know immediately how well they had done and would be assisting *E* as well. After making it clear that *S*s could keep their original lists, *E* instructed them to circle on the complete list those words they had constructed previously and to indicate in the space provided the number of words written. After the complete lists were collected, *S*s in the experimental groups were told that the reported reference group scores were fictitious and that performance on this task alone was not an adequate measure of creativity.

Results and Discussion

The *S*s who reported more words on the second day than they had written on the first were considered to have cheated; 59 (53%) of the 111 *S*s cheated. Since there were no sex dif-

ferences in incidence of cheating or for any variable discussed here, data from the two sexes are combined. As predicted, a greater proportion of *S*s in the Failure condition (61%) cheated than did those in the Control condition (43%; $CR = 1.49$, $p = .07$). Contrary to expectations, the incidence of cheating in the Success condition (56%) approximated that of the Failure condition but did not differ significantly from the Control condition.

The *S*s were assigned to High (H), Middle (M), and Low (L) anxiety levels on the basis of their scores on the Debilitating Anxiety Scale of the AAT. Fortunately for purposes of interpreting the cheating data, anxiety was not correlated with task performance. As predicted, anxiety scores proved to be positively associated with incidence of cheating: H vs. M, $CR = 2.30$, $p = .01$; H vs. L, $CR = 3.66$, $p < .001$; M vs. L, $CR = 1.49$, $p = .07$. Further analysis indicated that this effect holds only when knowledge of peer performance is provided; level of anxiety did not predict to cheating among subjects who received no information about peer performance. Although it was not taken into account in her interpretation, Gilligan's report of a positive relation between test anxiety and cheating was based on a procedure which involves the provision of knowledge of peer performance to every subject. The present study extends Gilligan's finding by demonstrating that the relation does not hold in the absence of knowledge about peer performance.

The present findings also indicate that peer knowledge induces cheating only among *S*s with moderate to high levels of anxiety. The effect of the peer knowledge manipulations on cheating differed at each of the three anxiety levels. Among *S*s with high anxiety, more cheating occurred in the Success and Failure conditions than in the Control condition (Success vs. Control, $CR = 1.82$, $p = .03$; Failure vs. Control, $CR = 1.45$, $p = .07$). Among moderately anxious *S*s, only the Failure vs. Control comparison proved significant ($CR = 1.61$, $p = .05$), whereas there were no differences between conditions among the *S*s with low anxiety.

Conclusions from studies which have found significant effects from peer knowledge manipulations (Hill & Kochendorfer, 1966; Shelton, 1966) need to be qualified in light of the present results. Such effects probably are attributable to *S*s with moderate to high anxiety since, in the present case, they are the *S*s who respond to the manipulations. Surprisingly, both Success and Failure elicit comparable levels of cheating from *S*s with high anxiety. Evidently knowledge of peer performance itself, and not an unfavorable discrepancy between peer performance and own performance, instigates cheating for the high-anxious *S*. This suggests that providing knowledge of peer performance to high-anxious *S*s may induce cheating because it is sufficient to cue an expectation of aversive consequences. That is, when high-anxious *S*s are placed in evaluative situations, it is assumed that they anticipate negative outcomes, and therefore aversive consequences, although available information suggests a more positive prognosis. Moderately anxious *S*s, on the other hand, appear to be influenced by the

informational value of peer knowledge; only unfavorable discrepancies between peer performance and own performance elicit cheating. The expectation of aversive consequences for failure is less generalized for the low-anxious *S*; unfavorable comparison with unknown others is not sufficient to induce cheating for him.

Further exploration of the range of applicability of our general hypothesis and this elaboration of it requires the implementation of a variety of peer knowledge manipulations. Both the manner in which such information is provided and the reference groups about which *S* is being informed should be varied. For example, the interactions between anxiety and peer knowledge obtained from our college-oriented *S*s may not hold when peer performance is presented as the attainment of a high proportion of the group as opposed to the average member. Such a sharpening of success and failure might elicit differential responding to these conditions by *S*s with high anxiety. On the other hand, low-anxious *S*s might be more influenced by peer performance information when it derived from the primary group rather than the more distant group employed here. Support for this suggestion comes from our additional finding that more *S*s with low than with high first day performance scores cheated ($CR = 1.85$, $p = .06$, two-tailed). This relation held within all anxiety levels. Interpreting this finding requires the assumption that social comparison took place after the first session. It can then be argued that low performance relative to that of the primary group instigated cheating among *S*s of all anxiety levels, in contrast to the absence of such an effect for low-anxious *S*s when information from a more distant group was provided.

References

Alpert, R., and Haber, R. M. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology,* 1960, **61**, 207–215.

Gilligan, C. F. Responses to temptation: An analysis of motives. Unpublished doctoral dissertation, Harvard University, 1963.

Hill, J. P., and Kochendorfer, R. A. Knowledge of peer scores and risk of detection as determinants of cheating in a resistance to temptation situation. Unpublished manuscript, University of Minnesota, 1966.

Shelton, L. G. The effects on cheating of knowledge of peer success and publicizing performance outcome to peers. Unpublished master's thesis, University of Minnesota, 1966.

3.5 *Alienation of Present-Day Adolescents*

LOUIS J. WISE, M.D.

Reprinted from the *Journal of the American Academy of Child Psychiatry,* 1970, **9,** 264–271, with permission of the author and the *Journal.* Louis J. Wise is a practicing child psychiatrist in Beverly Hills, California, and is also a clinical professor of psychiatry, University of Southern California School of Medicine.

Wise begins his paper by raising a number of questions about normal healthy adolescent behavior and psychopathological adolescent alienation, and uses a series of clinical vignettes to show the complexities involved in assessing the alienation process. The patients' feelings of detachment, rejection, and debasement were experienced as a sense of alienation, the treatment of which is complicated by the current fad to glamorize alienation. Wise also postulates that the more seriously disturbed adolescents had particularly faulty relationships with their mothers in the early years of life.

Youth today is pressed on many fronts, external and internal, having to live with the many complexities of current human concerns, problems, and relationships, and having to cope with and master the usual adolescent, psychological conflicts. In spite of the increased pressures, most adolescents, nevertheless, are continuing their schooling and striving for betterment in their lives, present and future. Many adolescents are taken with, and tempted by, the excitement of the drug scene and liberalized sexuality. However, the healthier adolescent, has, by virtue of his earlier life experiences, developed the capacity to integrate his impulses and to control them. He can come to terms with his sexuality while adopting a more mature and flexible superego; and above all, he can allow closeness and intimacy in human relationships while at the same time adopting a real sense of identity.

It becomes a greater and greater challenge to the psychiatric diagnostician to differentiate between essentially normal, healthy and psychopathological adolescent alienation, protest, rebellion, or dissent, and similar behavior. The fact that an adolescent wears long hair and a beard or hippie jewelry and attire, does not necessarily mean pathology. For many, it is just part of the scene. Is it to be considered pathological when one has turned on with marijuana and gets 'stoned' on this drug, or takes trips on LSD? If so, how frequently must these drugs be used and what is the degree of pathology to be assigned to this? Is rebellious behavior based on healthy self-assertiveness and self-realization in certain spheres, or is it pathologically pervasive in all relationships? Is a particular youth adhering to a cause or joining a fad as a means of escape, or as an exercise of integrity in maintaining an unpopular or misunderstood position? Can attempts at self-realization and self-identity be discerned in the esoteric interests of youth, as well as in their sometimes provocative strivings for ideals?

It is essential that we ask questions such as these, and that we attempt to answer them through a careful diagnostic appraisal—one which gives as much attention to evidence of healthy striving as to signs of pathology. The following clinical vignettes are offered as examples of the complexities involved in assessing the diagnostic and prognostic factors in the alienation process.

Jane

Seventeen-year-old Jane had done A and B work in high school well into her sophomore year. Her performance deteriorated to C's and D's in her senior year, at which time she was seen in psychotherapy. She had been using marijuana and had taken LSD 'trips' on a number of occasions. She had been involved in sexual affairs with a number of boys and was defiant in her behavior toward her parents and teachers. She considered that it was impossible to remain at home, and felt justified in doing as she pleased.

After several months of treatment, Jane described her feelings in relation to her behavior. "There was just a need

to be a part of a group and a need to feel secure. I guess I didn't feel that I could face up to being an individual and being different. I couldn't face being grown-up either. I was really in a bind. I wasn't about to face accepting new responsibilities, so I just plain alienated myself from the world and used drugs and behaved the way I did —and I didn't care—and yet, I must have. At least I know I worried. And doctor, as I look back to what I was doing during the last year or so, I must say I really felt worried about my behavior a good deal, even though I didn't admit it; even while I continued to protest and just had to make all that noise. Yet, I knew in my heart that something was wrong and that I wasn't adjusting the way I really wanted to do."

Jane, the younger of two daughters, had known she was adopted as an infant. Throughout her childhood, both parents had impressed on her how much they loved her and how they had "chosen" her, and she was literally stuffed with material tokens of affection. Physically attractive and obviously bright, Jane nevertheless could not accept herself as adequate or able to participate in close peer relationships. After discussing her estranged feelings, she expressed doubts about her origin and her adopted state. Eventually, it was discovered that her father was, indeed, her natural parent, a fact that her adoptive mother had not known. The father, who "arranged" the adoption even though he had himself fathered the child was responsible for the deception. Once the deception was uncovered and Jane made a party to the revelation, she accepted the therapeutic explanation that she felt in part alienated because she had reason to have doubted her father's integrity and, hence, his love for her. The defiance Jane demonstrated vis-a-vis adult authority, her use of drugs and sexual behavior, might have been understood as quite representative of current adolescent behavior; her slump in academic performance and negative attitude toward herself could also have been construed as evidence of protest and disillusionment. However, her repeated appeals for understanding and help, her complaints of loneliness and the need for security, suggested deeper disturbance.

As a baby, Jane was described by her mother as difficult to manage—fitful and cranky, a poor eater, slow in motor development, resistant to bowel and bladder training until late in her second year. Mother had been frequently absent due to several severe physical illnesses and major surgery during Jane's first two and a half years. Mother suggested she never did get to know her child, never could feel close to her and anxiously described not enjoying time spent with Jane, describing her as frequently sullen and quiet—"like she was depressed." Later, during the fourth to eighth year of life, Jane found it easy to turn to father who "was too wild, and excited Jane so," according to the mother. Since there appeared to be so much mutually needful and seductive behavior between daughter and father, Jane had difficulty in extricating herself from this relationship in order to individuate herself and to make ini-

tial efforts to separate herself from paternal ties.

Mary

Another girl of 15, described her reactions and thoughts about being part of the scene: "I was always scared I wouldn't be a success in life, according to the standards set by my parents and other adults. Everything seemed to be—get a college degree and make money. But what if one isn't interested in doing those things? What if you want to live in a different way? And then I'd think that they would laugh at me and I would wind up being a nobody. I hate to have those feelings of being lost and feeling empty. I can't help but feel resentful toward my parents. They never really got with it and never really were in tune with me."

This adolescent girl, the product of a professional father and a hypercritical and demanding mother, came into therapy at a time when she was panicked by academic demands at school and her own obsessive needs for perfection in her work. Terrified by possible failure, even though achieving quite well, she felt unloved, estranged from siblings and peers, and drove herself incessantly into bouts of intense anxiety, as she sought high performance levels to avoid her deep feelings of emerging rage at her parents and at herself. Lonely, lost, and empty, she eventually resolved her oedipal anxieties after confronting herself with her rage at being rejected by her father, even though she felt she was brighter than her mother and her two older siblings. Currently completing high school, she has turned to social activities, has accepted a less demanding academic performance for herself, and enjoys the possibilities of a future of marriage and of teaching children.

As an infant, she was reared and disciplined by a series of Negro maids, while mother helped father in his professional practice. A quiet child, moody, with occasional outbursts of rage when thwarted, she rarely seemed to respond positively to her mother's irregular and infrequent attempts to soothe or quiet her when troubled. Once grown into latency, she endured repeated frustrations in relation to her busy father.

Carol

After a period of intense self-examination of depressive feelings, during treatment, Carol described her 'sense of alienation' in the following manner: "Most of my sense of what one could call alienation, probably stems from the feeling that I always felt shut out, even from myself. I've always been the spectator, and never the participant, even in my own emotions. Why have I been excluded? I cannot describe my life as any but one of aesthetic gratification. I have grown up in a house that is lovely but cold, like a modish woman straining her clothes away from touch. My senses are full, but I don't believe in anything and I can only look down on my committed contemporaries. I look down on my protest-rally, draft-card burning, debate-club friends, and am filled with envy. I wish I were like them and could be with them. I've even tried to identify myself with several of my friends, but each time my personality

rejects the implant of this foreign tissue. I am terrified about the falsehoods and superficiality in people. Even in my different treatment attempts, I couldn't escape my own 'apartness,' and I decided to give them up rather than suffer the humiliation of asking for answers all the time. I only feel truly alive when with myself after some quite bruising experience. The shock of being hurt seems to assure me of my own fallibility, and thus of my membership in the human race. I often hurt myself almost consciously to jolt myself into a feeling which I consider essential. One of the most curiously wounding things I have come to realize is that people actually like me and I do like them in turn. I do have friends, in fact, many of them. Sometimes, I am confused by the very great, but undirected love I feel toward them. It is of two equally matched forces, my real emotions and the fear I have of them; and the articulation of this love is, therefore, doubly painful. I say, 'I love you,' and I am filled with despair. I feel as though I am struggling to reach an impossible goal. How can I expect to find peace with another person, when I am continually still at war with myself? Yet, I will not compromise one fraction of that embattled self. I am grateful for my friends, but they cannot help me. If I am 'alienated,' at least I am still free, and I will fight alone until satisfactory terms have been reached on both sides."

This bright and sensitive senior high school student had come into treatment because of trichotillomania (the eyebrows, lower eyelids and pubic hairs), bouts of depression, and an ambivalent relationship with both of her divorced parents. She recognized her parents' inability to love her, their egocentricity, and their continual demands for socially correct behavior and attitudes. Hurt and angered, Carol strove to lash back by intellectual attempts, utilizing dramatic skills in school performances, singing, and writing prose and poetry. She felt unfeminine, and constantly deprecated her attractiveness, picturing herself as ugly, mean and a "loner." Encouraged by several affairs with older boys, she, nevertheless, doubted their interest in her and her own adequacy.

Avoiding therapeutic attempts to uncover the dynamics of the trichotillomania which appeared related to her marked self-depreciative image of herself, Carol frustrated three previous psychotherapeutic attempts. "I just couldn't face recognizing the depth of my rage at myself—because I was so angry at my self-centered and very ungiving parents, especially my terribly mixed-up mother. She never did support or help me—I remember being so frequently hurt by my so-called friends—angered, I guess it was—and mother always appeared so cold and remote—she should never have had children—she was always trying to be a Dresden doll and didn't make that either. Then, of course, father—always so busy and ready to lecture with his pipe in hand, but really so remote, too—do you blame me for avoiding people, for fearing them, even my peers?"

Brad

Brad, a 19-year-old young man, who used marijuana and LSD to avoid reali-

ties, and to "float away in fantasy," reported: "I found myself so frequently and easily removed from people and my friends, and from all outer and once-familiar object situations and relationships. In particular, I felt alienated from myself, and this ultimately caused what I think of as a chain reaction, which eventually led to what they are writing about in the periodicals as 'alienation.' I thought of myself as being alienated even to the extent of changing to the point that I couldn't recognize myself as I once knew myself to be. It was being and feeling strange and very basically changed. Perhaps, the most major relationship wherein I felt this alienation is involved, had to do with my feelings about my relationship with my father. I felt that he never really loved me; I couldn't respect him because he wasn't a strong male figure. He always let mother intervene between us. I often had the feeling when I was a little boy that he wanted to come close to me and wanted to draw me close to him, but that he was afraid of arousing mother's resentment. Maybe this all had something to do with some latent feelings and tendencies I've always felt I had, where I might be wishing to seek approval and love from a male person, and perhaps this had to do with my wish to be close, and at the same time fearful of being close to other males as well as my father. So alienation for me on the whole is being removed or separated from the world. And then there was the so-called 'establishment' from which I really felt particularly alienated. This was easy for me to do because my father, being a successful businessman, was part of this establishment. So I would end up being pitted against all authority in schools, even my peers. So I wound up with a negative outlook on all strata of our society and on making it in our society, and I became upset."

Brad, the older of two teenage boys, remembered a peer-involved childhood, confirmed by both of his parents. The mother, however, later recalled her own depressive withdrawal from him during the three years after his birth. She stated, "I couldn't really pay much attention to Brad. Maybe I was lucky—the maid took over for me."

The father, a successful businessman, encouraged Brad as he grew into his teens to join him as a spectator in college and professional sports events. However, he, too, found it difficult to converse with Brad—"He would just turn me off—like that—I often thought Brad was in another world."

This alienated young man became acutely psychotic within several months after he offered the above comment about his own alienation.

Adolescents such as those described in the vignettes above reveal conflicts in peer relationships, trouble in relation to parents and other figures of authority, and, in spite of adequate intellectual capacities, have serious troubles in learning. These youngsters are moody and depressed, seem alone and isolated, and may use various drugs. Many, as a consequence of deep feelings of inadequacy, become very anxious and express fears of losing their minds, of having no directions or goals for themselves, of not knowing who they are or what they want to be, of feeling empty and lost.

Compared to other, more healthy, adolescents, these patients experience great difficulty in meeting and mastering the various tasks of adolescent development: emancipation from early infantile objects; the establishment of a mature sexual organization while replacing the childhood superego, which forbade sexual activity, with a more mature superego; developing adequate controls over emerging impulses; developing a capacity for productivity; resolving the struggle between dependency strivings and independent desires; achieving peer relationships and a capacity for physical and emotional intimacy; finding and accepting a firm and stable identity.

Theoretical Considerations

In my view, the adolescents described above have suffered severe and very early psychic traumata, primarily in relation to their mothers, from whom it became difficult to separate and individuate themselves. These were mothers who, for various reasons, seemed unable to provide the initial maternal care and relationship considered so necessary for that early, healthy symbiosis between mother and child, which later makes it possible for effective separation and healthy individuation in the child.

Mahler and Furer (1963), in evolving their concept of separation-individuation as a part of normal maturation and development, formulated that the infant, so to speak, undergoes a "hatching from the symbiotic membrane to become an individuated toddler." Building on this foundation, Blos (1967) suggested that adolescence in its totality can be viewed as "the second individuation process, the first one having been completed toward the end of the third year of life with the attainment of object constancy. Both periods have in common a heightened vulnerability of the personality organization. Both periods have in common the urgency for changes in psychic structure in consonance with the maturational forward surge. Last, but not least, both periods —should they miscarry—are followed by a specific deviant development (psychopathology) that embodies the respective failures of individuation" (p. 163).

As the infant attempts to individuate himself from his mother, he tends to experience a sense of loss, accompanied by what we understand as a depressive reaction. This leads to attempts on the child's part to reconstitute the relationship he once had with his mother. In this attempt, the mother functions as an "auxiliary ego" for the child (Spitz, 1965). Anna Freud (1954) has called this a "homeostatic equilibrium" which she conceptualizes as emanating from the symbiotic relationship of the child with the mother, and which, if fulfilling for both, should lead to the development of a healthy ego for the child. There are mothers who seem to force separation, threatening the infant with loss of love and care, while their unconscious tendencies are working against separation.

Rubinfine (1968), in discussing the child's early mastery of separateness from the mother, states, "When individuation occurs in normal development, there is no sudden reversal, no

abrupt withdrawal of love which gives rise to pathogenic amounts of aggression" (p. 408). The discovery that the mother is a separate being over whom one does not have omnipotent control is frequently accompanied by some depressive experience. Nevertheless, there is continued confidence in the loving characteristics of the object which makes it possible to continue to invest the object with positive good feelings even in its absence. However, many mothers themselves have conflicts about separation, and they seem to impose, in their attitudes and behavior, an inconsistency which makes it most difficult for the child to adapt to the separation.

Rubinfine suggests that a predisposition to depression originates in infancy, if there exists a rejecting and painful climate instead of one of warmth and gratification. As a consequence of an atmosphere of warmth and love in relation to his mother, the child initially invests the mother-child unity with narcissistic feelings. Later, when the differentiation of self and object is brought about, the child, aware of being separate and special, replaces the narcissistic union with object love.

Pumpian-Mindlin (1968), in discussing infantile omnipotence, sees the feeling of unlimited omnipotence, or "oceanic" feeling in the infant as representing the primary narcissistic need, which exists only so long as there is no concept of external objects. Ordinarily, with a mother present and involved with the infant, the real world comes to be acknowledged and ego formation begins. If the mother is physically or emotionally absent, unavoidable frustrations and inevitable delays in tension reduction must ensue. The existence of a real external world cannot be acknowledged. It is entirely conceivable that if an infant has an unstable or immature mother caring for him, one who herself may have separation and identity conflicts, he will have great difficulty in effecting separation from the original symbiotic tie to that mother. As a consequence, he becomes retarded in the process of individuating himself.

As a further point, Pumpian-Mindlin states that "with the realization of his own impotence and the attribution of omnipotence to the adults around him, the child comes to depend upon them for his self-esteem" (p. 9). One can speculate that such a situation in infancy might provide the beginnings of the low self-esteem seen in many an adolescent.

It is probable that many severely emotionally disturbed adolescents have had problems as indicated above in the early developmental relationships with their mothers. It would seem that they have never had an opportunity to work through the depressive affects during the separation-individuation phase, in part because their mothers were not able to help them work through this conflict.

Later, during the oedipal phase, the child is confronted with the necessity of giving up the parent of the opposite sex as the primary sexual object. Here, once again, the child finds himself frustrated. A mother with her own dependency problems may react with depressive affect in relation to the separation attempts by the child, as well as with attempts to sabotage developing relationships between father and child.

Again, the child fails to receive the necessary parental help in mastering the depressive affects and conflicts over his attempts to individuate himself from parental emotional ties. Many children who have had these experiences tend to be passive, ineffectual, clinging, and conforming as they go through their latency years. Such behavior tends to impede the development of the all-important peer associations, and the expansion of ego into healthy curiosity and interest in learning. Many latency-age children who have been confronted with this early developmental deficit have increasing difficulty in learning in school. They may be bright, but don't learn; they appear bored and disinterested.

As a consequence of the traumatic experiences experienced in the early conflict between mother and child, separation situations in the ensuing periods of development are usually accompanied by much anxiety and fearfulness. What these children seem to lack is a capacity to establish and maintain mutual and reciprocal, closely integrated relationships with others. Because of the earlier traumatic experiences, the individual, deep down, is wary of such closeness. He is in conflict between his wish for closeness and his understandable need to develop his own unique sense of individuation.

In a paper dealing with Erikson's concepts of intimacy and identity, Williams (1968) comments on the current social and familial forces which contribute to intimacy problems and alienation. He notes, in Western culture, the artificial separation between the individual's ego identity and his capacity for developing intimacy. He questions Erikson's socioculturally determined position on intimacy and identity that, "only after a reasonable sense of identity has been established, can real intimacy be possible." Williams stresses the lack of encouragement within our culture for further experiences with intimacy following the early mother-child symbiosis. He cites this lack as a major factor in the development of narcissistic, self-oriented "identities," which do not include the capacity for intimacy, and which contribute thereby to the alienation process in adolescence.

The author supports Blos' contention regarding adolescence, in that "both the task of disengagement from primary objects and the abandonment of infantile ego states, necessitate a return to early phases of development" (p. 185). However, this overwhelming regressive pull to infantile dependency gratification is altogether too tempting and, as a consequence, very frightening. What follows too often, at least in the case of the more disturbed youngsters, is marked detachment, rejection and debasement, which is subjectively experienced as a frightening and deep sense of alienation.

Guntrip (1962) holds that psychoanalysis has tried to understand the developing personality in terms of the "depressive position psychology." The clinical picture of the alienated youth described in this paper suggests that, because of the faulty mother-child relationship and the associated problems in separation and individuation, the infant so traumatized withdraws to avoid the continuing hurt and does indeed adopt the depressive position. Guntrip

goes on to suggest that the schizoid position is an even more important explanatory concept than the depressive position and that the depressive state is really a defense against the schizoid fears.

Although there may be differences of opinion as to which affect is defense against the other, this reporter suggests that, when the depressive affect is utilized and does not seem to work, then, in this alienated adolescent syndrome, schizoid defenses may be utilized. Since the parents too may be finding it difficult to face their depressive feelings around the loss of their adolescent, they might also turn to what may be considered schizoid postures — withdrawal, coldness, and rejection. A result may be continued estrangement for the adolescent; he feels antagonized and eventually alienated. He is in an emotional (identity) crisis, feeling, with some validity, that no one is available to support him and no one can emphatically understand him.

Perhaps this parent-adolescent conflict, which really is more common than heretofore recognized, might be understood as proceeding from both sides: the adolescent by virtue of psychological growth needs must separate himself and face the "loss" of his original emotional ties to his parents; the parents must acknowledge the legitimate emancipatory needs of their adolescent children and face their "loss." Depressive reactions for the adolescent and his parents are inevitable. Although the adolescent, in a sense, is rejecting the adults, he is in turn concerned about being rejected by them. The felt loss of love increases his feelings of worthlessness and guilt, which tend to engender depression. It is important to recognize that the alienated youth's attempt to individuate frequently is crushed when met by parental rejection.

Therapeutic Implications

One is impressed by the fact that few of the very alienated adolescents find their way into individual treatment. However, since those who are truly alienated see psychiatric treatment and the psychiatrist as part of the society that they both need and are trying to reject, one can understand why acceptance of treatment is so difficult. The writer's clinical experience indicates that some of those who feel or fear that they are verging in this direction do not really want to go all the way, and become sufficiently concerned about themselves to make themselves available for psychiatric treatment. Treatment for these youngsters must take into account the deficiencies mentioned, allowing considerable time for the establishment of the relationship between patient and therapist.

As psychiatrists and psychoanalysts, we are generally committed to an attitude which embodies clinical objectivity, relative uninvolvement and emotional distance. However when a patient is drifting toward isolation from human beings and affects, such distancing increases alienation. One must question whether we dare do this with these unhappy, lonely, uninvolved, uncommitted patients. On the other hand, attempts by therapists—who represent middle-

class society to these youngsters—to get too chummy may frighten them into a withdrawal from treatment.

The alienated adolescent is actually very resistant and refractory to treatment. Often apathetic and disinterested in the therapeutic goals held by the therapist, he tends to cling tenaciously to the peer group and its values. As Halleck (1967) indicates, a major therapeutic problem relates to the glamorization of alienation at the present time through the effects of the mass communication media and the arts. These focus almost exclusively upon the more extreme behavior patterns in society, and many have begun to see them as the norm. Alienation seems to have become ritualized and may represent power for the adolescent. The dropout seems to shake everyone up, and is a threat to the establishment's notions of desired teenage conformity. Treatment of these youngsters must allow for flexibility and deviation from conventional treatment approaches. According to Halleck, one must recognize the disparity between the patient's verbalizations and protestations and his actual maladaptive behavior. As therapists, we must participate more in the relationship and be less a "blank screen": we have to be willing to set limits about lateness, missed hours, and other kinds of destructive or self-destructive attitudes and behavior; we must constantly guard against becoming impatient and against other countertransference reactions, especially so since we are confronted with acting-out behavior, the use of drugs, indiscriminate sexual attitudes and behavior, and conflicts in relation to various forms of authority and school work. In addition, the use of family sessions, with or without the patient, may be especially useful in educating parents about the nature of the problems that exist. Group therapy has been recommended by some as a successful adjunct to the individual therapy sessions.

Summary and Conclusions

This paper attempts to illustrate and delinate adolescent alienation, rebellion, and dissent. It suggests that developmental theory allows us to make helpful formulations regarding the genetic and dynamic forces contributing to the alienation of our youngsters.

It is postulated that the more seriously emotionally disturbed adolescents who alienate themselves from society have been seriously traumatized in the relationship with their mothers in the first years of life. It appears that progression from the symbolic relationship into and through the phase of separation and individuation was seriously impaired and that adequate mastery of those depressive affects inherent in the process of human separation was blocked.

An effort has been made to show how the alienated adolescent and his parents respond to one another—how each rejects the other. Suggestions are offered for various approaches to therapeutic interventions, including individual psychotherapy and conjoint family or group meetings, the latter mainly in attempting to clarify and resolve familial resistance to change.

References

Blos, P. (1967), The second individuation process of adolescence. *The Psychoanalytic Study of the Child,* **22**:162–186. New York: International Universities Press.

Freud, A. (1954), quoted by Margaret S. Mahler in "Problems of infantile neurosis: a discussion." *The Psychoanalytic Study of the Child,* **9**:65–66. New York: International Universities Press.

Guntrip, H. (1962), The manic-depressive problem in the light of the schizoid process. *Int. J. Psycho-Anal.,* **43**:98–112.

Halleck, S. L. (1967), Psychiatric treatment of alienated college students. *Amer. J. Psychiat.,* **124**:642–650.

Mahler, M. S. and Furer, M. (1963), Certain aspects of the separation-individuation phase. *Psychoanal. Quart.,* **32**:1–14.

Pumpian-Mindlin, E. (1968), Vicissitudes of Infantile Omnipotence. Presented at New Orleans Psychoanalytic Institute.

Rubinfine, D. L. (1968), Notes on a theory of depression. *Psychoanal. Quart.* **37**:400–417.

Spitz, R. A. (1965), *The First Year of Life.* New York: International Universities Press.

Williams, F. S. (1968), Erik Erikson's Concepts of Identity and Intimacy: Reviewed in Light of Present-Day Youth Alienation. Paper presented to Southern California Society for Adolescent Psychiatry, Los Angeles, California.

3.6 *Reducing Educational Pressures*

RICHARD DEMPSEY

Reprinted from *Science,* 1967, **147**, 1117, with permission of the author and the American Association for the Advancement of Science. Copyright © 1967 by the American Association for the Advancement of Science. Richard Dempsey is an administrator with the public schools of Darien, Connecticut, and is now Chairman, Department of Secondary Education of the University of Connecticut at Storrs.

Dempsey shows three ways in which negative pressures were reduced in a high school, resulting in a more relaxed, better-prepared student. However, he asks us to recognize that internal pressures are embedded in the individual and cannot be readily dissipated by outsiders. Furthermore, he reminds us that what constitutes a pressure for one child can be a pleasure and stimulation for another. Not all pressures are bad—some children thrive on them.

Few people can adequately define pressure in terms which embrace all youth. Although we commonly think of pressure as being entirely negative, that pressure (or drive, or need, or value orientation that motivates) to achieve can be favorable if it results in an action that has hope of a satisfying culmination. But when pressure cannot be translated into positive action, it results in frustration and anxiety. The homework assignment that requires 2 hours of diligent concentration may be an undesirable force for one child and a positive force for another. Furthermore, a day or a week hence, the same assignment may elicit reverse pressure on the two youths. What constitutes a pressure for one child does not necessarily constitute a pressure for another. Furthermore, not all pressures are universally evil—some youths thrive upon pressure.

One should point out that pressure can come from external and internal sources. The school, the teacher, and the parent have a degree of control over a portion of the external causes of pressure. They can regulate the curriculum or the assignment or the social obligations if they wish. Internal pressures, however, are more deep-rooted and cannot be allayed as easily. How a child perceives a given situation, based upon his own set of values and experiences, is generally controlled by the individual alone. The degree to which he displays psychic tension prior to taking college board examinations is embedded in the individual and cannot be readily dissipated by outsiders. In short, perhaps adults can do something about some of the pressures facing youth, but we must realize that we cannot mitigate all of them. During the past 4 years I have been director of a high school which was deeply concerned about pressure on youth. The faculty undertook to reduce what it believed to be negative pressures.

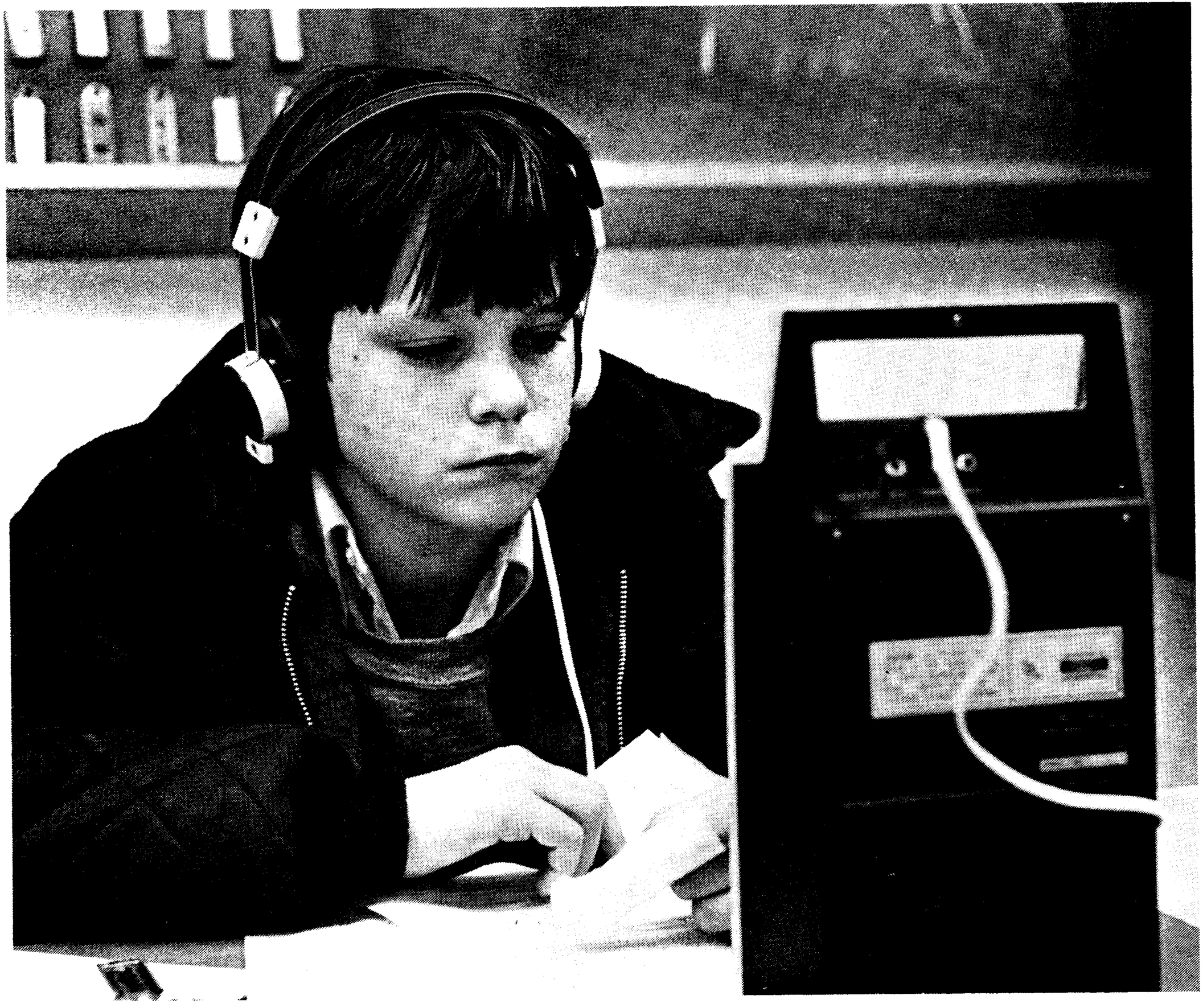
First, the schedule was changed from the traditional eight periods of 55 minutes to three 110-minute periods daily. Students meet with teachers on alternate days and once each week all classes meet—an English class might meet on Monday for one 55-minute period and on Wednesday and Friday for double periods of 110 minutes. Furthermore, faculty are compelled to provide a "break" sometime during each session to lessen the possible strain of sitting for a 2-hour lecture. This unconventional schedule has elicited very positive reaction from students, parents, and teachers. Within the schedule, students are able to carry more courses with no deleterious effect on either their tested achievement or grade-point averages. Longer class periods for discussion and supervised study, less frequent assignments, and greater variety of presentation increased the students' confidence in their ability to perform well.

Second, the athletic program was altered. A new coaching staff was recruited of men who were more interested in the general welfare of the student than in winning records. (These men have been known to bench a player who was not doing his best in the academic program.) Also, the number and frequency of athletic contests was reduced to eliminate undue tension buildup among both the partici-

pants and the observers. Athletes were not permitted to participate in more than one sport during a given season—a common occurrence in small schools. The reduced emphasis on athletics seems to have produced a more relaxed athlete, yet resulted in a program which now encompasses 15 percent of the males and produced winning teams in every sport.

A third approach to relieving pressure was through a philosophy called Idea—Interdisciplinary Education Approach. In essence this is merely a fusion of like subjects which are taught by a team of teachers. American literature, American history, American art, and American music are fused into a course called American studies which meets 2 hours daily. Several other common areas were similarly combined and the Chemical Education Material Study, the Physical Science Study Committee, and math programs are scheduled for fusion in the fall and will provide R & D in the industrial arts, thus serving as a practical application of these combined studies. Such combined programming has reduced the need for students to pull together isolated threads of knowledge and enabled them to see the broader aspect of the educational picture. It is also hoped that the combined modern sciences will allow the instructors to use their ingenuity and creativity and the students to apply their own innovative talents to practical problems. Thus the student will no longer see biology, chemistry, physics, or mathematics in isolation but in a harmoniously woven program.

I believe that the institution of a few external alterations in the school's overall program has resulted in a more relaxed and better prepared student. To support this, a marked change has occurred in the level of students' anxieties. The data obtained from the Taylor Manifest Anxiety Scale between 1963 and the present indicates a reduction in the anxiety level of students from a frighteningly high mean score of 33.2 to a comfortable 15.9. Furthermore, these changes may have been responsible for less vandalism among students, less truancy, and less internal strife among cliques, all of which cannot be substantiated statistically but have been observed by the faculty and administration of the school. It is my feeling that when parents, teachers, and students can work together as a group to allay what they feel are necessary pressures on youth, much can be done to remove the negative external stimuli on students.



Sybil Shackman from Monkmeyer

PART 4

The Learning Process: Theory, Research, and Practice

4.1 *Psychoneuro-biochemeducation*

DAVID KRECH

Reprinted from *Phi Delta Kappan,* 1969, **50**, 370–375, with permission of the author and *Phi Delta Kappa,* Inc. David Krech is on the faculty of the University of California at Berkeley. His research and writing have won him an international reputation in experimental, physiological, social, and personality psychology.

Krech's paper consists of an address to the American Association of School Administrators, in which he reviews a vast amount of research covering early experience, environmental complexity, cognitive development, and brain chemistry and speculates about its relevance for education and educators. The burden of his message is summarized in his contention that "Educators probably change brain structure and chemistry to a greater degree than any biochemist in the business."

The discovery that learning produces structural and chemical changes shows that it is not a temporary and transitory experience; indeed, the fact that learning takes place at all is as significant as what is learned and how it is learned.

I am a rat-brain psychologist with a weakness for speculation. Now time was when rat research was a fairly harmless activity, pursued by underpaid, dedicated, well-meaning characters. The world took little note and cared even less about our researches on how rats learned to thread their way through mazes. Oh, occasionally a misguided educator would take us seriously and try to fashion an educational psychology out of our rats-in-a-maze studies. But the classroom teachers—once removed from the school of education—would quickly see through such nonsense, and, forsaking all rats, would turn to the serious and difficult task of teaching children—unencumbered and unaided by our research and theory.

But time no longer is. Our psychology —especially when combined with educational practice and theory—must now be listed among the Powerful and, even perhaps, the Dangerous sciences. I refer specifically to the recent research developments in brain biochemistry and behavior—to some of which research I now turn.

The research I will discuss really concerns itself with the venerable mind-body problem beloved of philosophers and theologians. For brain biochemistry and behavior research seeks to find the *physical* basis for memory. In essence it asks the following question: In what corporal forms do we retain the remembrance of things past? What are the chemical or neurological or anatomical substrates of the evocative ghosts we call "memories"? Over the centuries of thought and decades of scientific research we have gained but very little on this question. Today, however, there is a feeling abroad that we are on the verge of great discoveries. Indeed, some researchers believe that we already know, in the rough, the form the final answer will take to the question I have raised. And it is this: The physical basis of any memory, whatever else it may be, involves either the production of new proteins, the release of differentiated molecules of ribonucleic acids (RNA's) or the induction of higher enzymatic activity levels in the brain. In a word, for every separate memory in the mind we will eventually find a differentiated chemical in the brain—"chemical memory pellets," as it were.

What warrant do we have for such a prophecy? To begin with, we have reason to believe that the storage of memory in the brain is a many-splendored, multi-phased, actively changing affair. That is, any single memory is not merely "deposited" in a completed form in the brain. Rather, it goes through a complex developmental history in the brain in which it changes from a short-term into a long-term memory. And each stage in this consolidation process seems to be dependent upon different although interrelated chemical mechanisms. Let me indicate to you one set (of quite a number which are now available) of speculative hypotheses concerning this developmental transformation of memories.

First we can assume that immediately after every experience, a relatively short-lived reverberatory process is set up within the brain. This process continues for a time after the stimulus disappears and permits us to remember

events which occurred moments or minutes ago. But this reverberatory process fairly quickly decays and disappears—and as it does, so does the related memory. However, under certain conditions, the short-term reverberatory process, before it disappears completely from the scene, triggers off a second and quite different series of events in the brain. This second series of events involves the release of new RNA's or the production of new proteins and other macromolecules. And these chemical changes are relatively long-lasting and serve as the physical bases of our long-term memories.

Now it can be supposed that if we increased the robustness or the survival time of the initial reverberatory process we might increase the probability of converting the short-term memory into a long-term memory. There are several ways one could do that. Through the repetition of the same stimulus one could presumably prolong or continually reinstate the reverberatory process and thus, perhaps, make it more effective in inducing permanent chemical changes in the brain. The old-fashioned term for this procedure is "drill" or "practice," and drill and practice are indeed effective techniques for helping the conversion of short-term memories into long-term ones.

But James McGaugh, at the University of California at Irvine, got the bright idea that he could achieve much the same results chemically. His argument —very much simplified—went something like this: A drug which would increase neural and chemical activity within the brain might either increase the vigor of the reverberatory process, or the ease with which the long-term chemical processes would "take off," and thus facilitate the conversion of short-term memories into long-term ones. Apparently his idea was a sound one, for with the use of chemical compounds like strychnine and metrazol, which are central nervous system stimulants, McGaugh has been eminently successful in raising the intellectual level of hundreds of southern California mice.

In one of his experiments which is most pregnant with social implications and promises and forebodings for the future, McGaugh tested the maze-learning ability of two quite different strains of mice. One of the strains was, by heredity, particularly adept at maze learning; the other, particularly stupid at that task. Some animals from each strain were injected with different doses of metrazol after each daily learning trial to see whether there would be an improvement in their ability to retain what they had learned on that trial—and some were not. The findings pleased everyone—presumably even the mice. With the optimal dosage of metrazol, the chemically treated mice were 40 percent better in remembering their daily lessons than were their untreated brothers. Indeed, under metrazol treatment the hereditarily stupid mice were able to turn in better performances than their hereditarily superior but untreated colleagues. Here we have a "chemical memory pill" which not only improves memory and learning but can serve to make all mice equal whom God—or genetics—hath created unequal. May I suggest that some place

in the back of your mind, you might begin to speculate on what it can mean —socially, educationally, politically—if and when we find drugs which will be similarly effective for human beings.

But let me continue with my story. What chemistry can give, it can also take away—as Agranoff and his now notorious goldfish at the University of Michigan have shown. Agranoff argued that if we could prevent the brain from manufacturing the chemicals involved in the long-term memory process, then we would create an animal which might have normal short-term memories, but would be incapable of establishing enduring memories. Agranoff trained his fish to swim from one side of an aquarium to another, whenever a signal light was turned on, in order to avoid an electric shock. Goldfish can learn this task within a 40-minute period, and once it is learned, they remember it over many days. Now Agranoff varied his experiments. Immediately before, and in some experiments immediately after, training, Agranoff injected puromycin or actinomycin-D (two antibiotics which prevent the formation of new proteins or nuclear RNA) into the brains of a new group of goldfish. His findings were most encouraging (to Agranoff, that is, not necessarily to the goldfish). The injected goldfish were not impaired in their *learning* of the shock-avoidance task since, presumably, the short-term reverberatory process which enables a fish to remember its lesson from one trial to another—a matter of a few seconds—does not involve the synthesis of new proteins or nuclear RNA. But when tested a day or two later the fish showed almost no retention for the task they had known so well the day before—indicating that the long-term process *is* dependent upon the synthesis of these compounds in the brain. Here, then, we find not only support for our general theory but we have a suggestion that there exist in antimetabolites whole families of chemical memory preventatives which seem not to interfere with the individual's immediate capacity to obey immediate orders, but which do prevent him from building up a permanent body of experiences, expectations, and skills. Conjure up, if you are of that mind, what evils such weapons can wreak in the hands of the Orwellian authorities of 1984—but I must hurry on to our next set of experiments.

A number of years ago, James McConnell at the University of Michigan threw all the brain researchers into a tizzy by reporting that he had succeeded in teaching planaria—a fairly primitive type of flatworm—to make a simple response to a light signal, that he then ground up his educated flatworms, fed the pieces to untrained fellow worms—and lo and behold, the uneducated flatworms wound up with the *memories* of the worms which they had just eaten, and, without any training, could perform the response of the late-lamented and digested "donor" worms!

But then all hell broke loose when other workers in other laboratories and in other countries reported that they could train a *rat,* make an extract from its brain, inject this extract into an untrained rat, and by so doing cause the recipient rat to acquire the memories of the now-dead donor rat. It is one thing to claim this for the primitive planaria,

which, after all, do not have very much in the way of a structurally differentiated and organized brain. It is a very different thing to claim it for the rat, which *is* a serious mammal, with a highly developed brain, not too different in complexity, in differentiation, and in organization from our own.

The dust raised by these reports has not yet settled. Indeed, most scientists are definitely on the side of the non-believers—but the work goes on, and we cannot predict the final outcome of these experiments, many of which have given negative results. However, as a result of this work, a number of brain researchers have been moved, over the last two or three years, from the position of stiff-necked disbelief to the position of "well, maybe—I don't believe it, but well, maybe." And this is where *I* stand at the moment—fearless and foursquare proclaiming "well, maybe. . . ." Now, if it should come to pass that McConnell and his fellow believers are right, then we will indeed have made a huge jump forward. For we would then have a most effective behavioral assay method which should enable us to zero in on this marvelous brain-goulash which can transfer information from one brain to another, and isolate and identify in detail all the "memory" proteins, enzymes, RNA's, or other macromolecules. After that—the world of the mind is ours! But that day is not here yet. Let me leave these brave new world experimenters and go on with another question and another set of experiments.

Does the research I have reviewed mean that if and when we will have developed get-smart pills (*a la* McGaugh), or chemical erasures of wrong mental habits (*a la* Agranoff), or specific knowledge pills (*a la* McConnell), we will be able to do without Head Start programs, educational enrichment programs, school supervisors, educational research, and, indeed, without most of our educational paraphernalia? The answer to this question, gentlemen, is a most reassuring "NO." I might even say, *"Au contraire."* Precisely because of the advances in brain biochemistry, the significance of the educator will be greatly increased—*and just as greatly changed.* Let me tell you why I think so by describing to you the results of some of our own work in the Berkeley laboratories.

Some time ago we set ourselves the following problem: If the laying down of memories involves the synthesis of chemical products in the brain, then one should find that an animal which has lived a life replete with opportunities for learning and memorizing would end with a brain chemically and morphologically different from an animal which has lived out an intellectually impoverished life. For almost two decades, now, E. L. Bennett, Marion Diamond, M. R. Rosenzweig, and I, together with technical assistants, graduate students, and thousands of rats, have labored—and some of us have even sacrificed our lives—to find such evidence. Let me tell you some of what we found.

At weaning time we divide our experimental rats into two groups, half of the rats being placed in an "intellectually enriched" environment, the other half—their brothers—in the deprived environment. While both groups receive

identical food and water, their psychological environments differ greatly. The animals in the first group live together in one large cage, are provided with many rat toys (tunnels to explore, ladders to climb, levers to press), and they are assigned to graduate students who are admonished to give these rats loving care and kindness, teach them to run mazes, and in general to provide them with the best and most expensive supervised higher education available to any young rat at the University of California. While these rats are thus being encouraged to store up many and varied memories, their brother rats, in the deprived group, live in isolated, barren cages, devoid of stimulation by either their environmental appurtenances, fellow rats, or graduate students. After about 80 days of this differential treatment, all the animals are sacrificed, their brains dissected out and various chemical and histological analyses performed. The results are convincing. The brain from a rat from the enriched environment—and presumably, therefore, with many more stored memories—has a heavier and thicker cortex, a better blood supply, larger brain cells, more glia cells, and increased activity of two brain enzymes, acetylcholinesterase and cholinesterase, than does the brain from an animal whose life has been less memorable.

We can draw several morals from these experiments. First, the growing animal's psychological environment is of crucial importance for the development of its brain. By manipulating the environment of the young, one can truly create a "lame brain"—with lighter cortex, shrunken brain cells, fewer glia cells, smaller blood vessels, and lower enzymatic activity levels—or one can create a more robust, a healthier, a more metabolically active brain. If it should turn out that what is true for the rat brain is also true for the human brain, and that by careful manipulation of this or that group's early environment we can develop among them bigger and better brains or smaller and meaner ones, the wondrous promises of a glorious future or the monstrous horrors of a Huxlian brave new world are fairly self-evident.

The second conclusion I draw from our experiments is this: Since the effect of any chemical upon an organ is, in part, a function of the beginning chemical status of that organ, and since—as we have just seen—the chemical and anatomical status of the individual's brain is determined by his educational experience, then the effectiveness of the biochemist's "get smart pill" will depend upon how the educator has prepared the brain in the first instance. Indeed, a review of all the data indicates that manipulating the educational and psychological environment is a more effective way of inducing long-lasting brain changes than direct administration of drugs. Educators probably change brain structure and chemistry to a greater degree than any biochemist in the business. Another way of saying this is: The educator *can potentiate or undo the work of the brain biochemist.*

But there is still more to report, and more lessons to draw. Consider the experimental problem we faced when we tried to create a psychologically enriched environment for our Berkeley rats. We did not really know how, so

we threw everything into the environment, including, almost the kitchen sink, and called it "a psychologically enriched environment." The cages were kept in brightly lighted, sound-filled rooms; the rats were given playmates to relate to, games to manipulate, maze problems to solve, new areas to explore. They were fondled and tamed and chucked under the chin at the drop of a site-visitor. In other words, we provided our happy rats with almost every kind of stimulation we could think of—or afford. And it seems to have worked. But of course it is quite possible that in our "kitchen-sink design," many of the things we did were not at all necessary—indeed, some may have had an adverse effect. And so we undertook a series of experiments to discover which elements of our environment were effective and which were not. I shall not bore you with the details of the many experiments already run and the many more which are now being run in the Berkeley laboratory. Let me list, however, some of the tentative conclusions which one can already make:

First. Sheer exercise or physical activity alone is not at all effective in developing the brain. A physical training director seems not to be an adequate substitute for a teacher.

Second. Varied visual stimulation, or indeed any kind of visual stimulation, is neither necessary nor sufficient to develop the brain, as we were able to demonstrate by using rats blinded at weaning age.

Third. Handling, or taming, or petting is also without effect in developing the growing rat's brain. Love is Not Enough.

Fourth. The presence of a brother rat in our intellectually deprived rat's cage helps him not a whit. *Bruderschaft* is not enough.

Fifth. Teaching the rat to press levers for food—that and only that seems to help somewhat, but only minimally. Not every problem-set will do, either.

The only experience we have thus far found really effective is freedom to roam around in a large object-filled space. From a recent experiment in Diamond's laboratory there are some suggestions that if the young rat is given continuous and varied maze-problems to solve—that and little else—the rat will develop a number of the same brain changes (at least the morphological ones) which we had observed in our randomly "enriched" environment.

It is clear, then, that not *every* experience or variation in stimulation contributes equally to the development of the brain. But of even greater interest is the suggestion in the above data that the most effective way to develop the brain is through what I will call *species-specific enrichment experiences.* Here is what I mean: The ability of a rat to learn its way through tunnels and dark passages, to localize points in a three-dimensional space full of objects to be climbed upon, burrowed under, and crawled through is, we can assume, of particular survival value for the rat as he is now constituted. Presumably, through the selective evolutionary process, the rat has developed a brain which is

peculiarly fitted to support and enhance these skills. The "effective rat brain," therefore, is one which is a good "space-brain"—not a lever-pressing brain or an arithmetic-reasoning brain. The effective stimulating environment, correspondingly, would be one which makes *spatial learning* demands on that brain—which "pushes" that particular kind of brain in that particular way. To generalize this hypothesis, I would suggest that *for each species there exists a set of species-specific experiences which are maximally enriching and which are maximally efficient in developing its brain.*

If there be any validity to my hypothesis, then the challenge to the human educator is clear. For the educator, too, you may have noticed, has been using the kitchen-sink approach when he seeks to design a psychologically or educationally enriched environment for the child. Some educators would bombard the child—practically from infancy on—with every kind of stimulus change imaginable. His crib is festooned with jumping beads and dangling colored bits and pieces of wood (all sold very expensively to his affluent parents); he is given squishy, squeaking, squawking toys to play with, to fondle, to be frightened by, to choke on. He is jounced and bounced and picked up and put down. And when he goes to school—he finds the same blooming, buzzing confusion. He is stimulated with play activities, with opportunities for social interaction, with rhythmic movements, with music, with visual displays, with contact sports, with tactual experiences, and with anything and everything which the school system can think of—or afford. But it may be that a "stimulating environment" and an "enriched environment" are not one and the same thing. It is not true that a brain is a brain is a brain. The rat is a rat and he hath a rat's brain; the child is a child and he hath a child's brain—and each, according to my hypothesis, requires its own educational nutrient. What, then, are the species-specific enrichments for the human child?

Of course I do not know the answer to this question, but let me share with you my present enthusiastic guess that in the language arts will you find part of the answer.

I can start with no better text than a quotation from my teacher, Edward Chace Tolman, who was a completely devoted rat psychologist. "Speech," he wrote, ". . . is in any really developed and characteristic sense, the sole prerogative of the human being. . . . It is speech which first and foremost distinguishes man from the great apes." (1932)[1] In my opinion, it is in the study of language, above anything else, that the psychologist will discover the psychology of man, and that the educator will discover how to educate man.

In the first place, and we must be clear about this, human language, with its complex and *abstract structure,* has *nothing* in common with animal communication. Language is probably the clearest instance of a pure species-specific behavior. This is true whether you study language as a neurologist, or as a psychologist. Let us look at some brain research first.

Recently Robinson, at the National Institute of Mental Health (1967), at-

tempted to discover which areas of the monkey's brain controlled its vocalizations.[2] Now the monkey most certainly uses vocalization for communication, but principally for communications with emotional tone such as threat, fear, pain, and pleasure. In Robinson's study 15 unanesthetized animals, with brains exposed by surgery, were used. Some 5,880 different loci or spots in the brain were stimulated by electrodes to see whether such stimulation could bring forth vocalization. The loci explored included neocortical areas as well as areas in the limbic system, that older part of the mammalian brain which is most intimately involved with motivational and emotional responses.

Robinson's results were clear-cut: First, despite his exploration of several hundred different neocortical sites he was unable to raise a single sound from his animals by stimulating their *neocortex.* Second, stimulation of the limbic system brought forth regular, consistent, and identifiable vocalizations.

These results differ sharply from those found with the human brain. While there is some evidence that human cries and exclamations—uttered in moments of excitement—are also controlled by the limbic system, *speech and language clearly depend upon neocortical areas*—areas for which there simply are no analogues in the brain of any other animal. These areas are, of course, the well-known Broca and Wernicke areas in the left hemisphere of the human brain. It seems clear, as Robinson puts it, that "human speech did not develop out of primate vocalization, but arose from *new tissue* [italics my own] which permitted it the necessary detachment from immediate, emotional situations." Man's brain, *and man's brain alone,* is a language-supporting brain.

Corresponding to the neurological picture is the psycholinguist's view of language. Almost every psycholinguist is impressed not only with the unique nature of language itself but with its unique mode of achievement by the child. Whatever value so-called reinforcement or stimulus-response theories of learning may have for describing acquisition of motor skills by people, maze-learning by rats, and bar-pressing by pigeons—these theories are assessed as completely trivial and utterly irrelevant when it comes to understanding that "stunning intellectual achievement" (McNeill, 1966[3]), the acquisition of language by the child. Indeed, in reading the psycholinguist's work one is left with the impression that we will have to develop a species-specific learning theory for this species-specific behavior of language. I must confess that I agree with them. And if we ever achieve an understanding of language development, and if we learn how to push the *human* brain with this *human* experience, then will we indeed be on our way.

I know that other people have proposed other ways with which to enrich the child's education. Some plug for what are referred to as "cognitive" experience or "productive thinking" experiences, etc. Let me hasten to record that I quite agree with them. As a matter of fact, I am not at all certain that I am saying anything other than what

my cognitive friends propose. For I hold with McNeill's judgment that ". . . the study of how language is acquired may provide insight into the very basis of mental life." And, I would go on, being human *means* having an effective mental, cognitive life.

It is for these and many, many other reasons that I would urge the educator to turn to the psycholinguist—as well as to Piaget and Crutchfield and Bruner —for his major guides in designing a rational educational enrichment program.

Whether my guess merits this enthusiasm or not will perhaps eventually be determined by research. But here is the challenge and here is the promise for the educator. Drop your kitchen-sink approach, and specify and define for us the species-specific psychologically enriching experiences for the child—and we will be off and running!

Where will we run? Let me speculate out loud. It is perfectly reasonable to suppose that we will be able to find specific biochemical boosters and biochemical inhibitors for different kinds of memories and imagery, or for different kinds of abilities, or for different kinds of personality or temperament traits. With such chemical agents in hand, and with appropriate educational and training procedures, we may use them as supplementary therapy for those failing in this or that trait and thus will we be able to rectify and heal some of the mentally retarded and the senile. Of course we may use these agents for evil—to create docile, intellectually limited, but efficient human beasts of burden without memories beyond the order of the day (remember Agranoff's fish?).

But above all, there will be great changes made in the first and foremost and continuing business of society: the education and training of the young. The development of the mind of the child will come to rest in the knowledge and skills of the biochemist, and pharmacologist, and neurologist, and psychologist, and educator. And there will be a new expert abroad in the land — the psychoneurobiochemeducator. This multi-hybrid expert will have recourse—as I have suggested elsewhere—to protein memory consolidators, antimetabolite memory inhibitors, enzymatic learning stimulants, and many other potions and elixers of the mind from our new psychoneurobiochemopharmacopia.

There is a grievous problem here, however. Experts, whatever else they may be, are notorious ordertakers. *Who* will direct our psychoneurobiochemeducator where to work his expertise, and *what* shall we tell him to do? Here we are talking about goals, values, and aims. Shall our expert raise or lower docility, aggressiveness, musical ability, engineering ability, artistic sensitivity, effective intellectual functioning? Shall different ethnic or racial or national or social groups receive different treatments? In past centuries, and even today, this differential group treatment is precisely what our relatively primitive but quite effective medical and educational experts have been ordered by us to carry out. And lo, they have done so! On one side of the town they have created enclaves of

the sickly, the weak, the ignorant, the unskilled—in a word, the brutalized social vanquished. On the other side of the town they have created the social victors—the healthy, the strong, the knowledgeable, the skilled. Will we continue to do this in the future with our much more sophisticated and effective psychoneurobiochemeducators? Who, in other words, will control the brain controllers—and to what ends?

I have thought and worried about these questions, and I must confess to you that I cannot avoid a dread feeling of unease about the future.

At the same time I keep whistling the following tune in an attempt to cheer myself up: If there be any validity at all to my speculations this afternoon, they add up to this: The biochemist, neurologist, psychologist, and educator will eventually add to the intellectual stature of man. With this in mind, and clinging to a life-long faith in the virtues of knowledge and the intellect (for certainly, at this stage I can do no less), I find myself believing that man who by taking thought will have added cubits to his intellectual stature, will also acquire the added bit of wisdom and humaneness that will save us all. Let me stop on this note—before I scrutinize this faith and this hope too carefully.

References

1. Edward Chace Tolman, *Purposive Behavior in Animals and Men.* New York: The Century Company, 1932.
2. B. W. Robinson, "Vocalization Evoked from Forebrain in *Macaca Mulatta,*" *Physiology and Behavior,* 1967, No. 2, pp. 345–54.
3. D. McNeill, "The Creation of Language," *Discovery,* 1966, No. 27, pp. 34–38.

4.2 *A Self-Concept Theory of Learning: A Learning Theory for Teachers*

WALCOTT H. BEATTY AND RODNEY CLARK

Reprinted with permission of the authors from a report prepared in 1961 (revised in 1962) for the Teacher Education Project of San Francisco State College, an undertaking supported by a grant from the National Institute of Mental Health. Both authors are psychologists who are interested in the application of self theory and principles of group process to learning problems encountered in the classroom as well as in the world outside the school. Walcott H. Beatty is professor of psychology, and Rodney Clark is professor of education at San Francisco State College.

Most learning problems, as teachers know well, are directly or indirectly involved with motivation; hence an understanding of the learner's motivation is therefore essential if we are to help him progress. Motivation is related to the attitudes, feelings, and emotions the learner has with respect to learning tasks. In order to understand this affective dimension of the learner, in turn, we must empathize with him—that is, we must develop some awareness of how he regards himself and his world.

A Learning Theory for Teachers

The development and improvement of education depends on a sound and seminal theory of learning, directly applicable to classrooms. The needed theory must comprehend all aspects of the relations between the school as a social institution and the development of the human beings who are part of it. A learning theory for teachers should make it possible for them to consider the live students with whom they work in the specific experiences provided by current curricula.

A learning theory which seriously attempts to integrate accepted psychological assumptions and concepts, and to contradict no known fact, may be widely at variance with much that teachers and students do in school today. Present practices have been derived from a variety of sources and most of them have not been the subject of careful research. As Estes points out in his review of learning in the *Encyclopedia of Educational Research* (4), there is a clear need for harmonization between the laboratory and classroom approaches to human learning. A useful theory must grow from verified knowledge about people and how they learn, no matter how that challenges present practices. In fact, the very challenge should stir imaginative explorations leading to new knowledge and thereby lay the foundations for more effective educational effort. It is hoped that this paper may serve as a beginning in that direction.

Three Assumptions

The things we do with people and the theories we propound to explain their behavior are based upon certain fundamental assumptions as to the nature of man. Thinking within the fields of psychology and psychiatry has changed radically in regard to what these assumptions are. We have learned a great deal about biological and psychological functioning, and as a result we have switched from mechanical explanations of man's behavior to more dynamic accounts. Psychologists with varied interests have been converging on a common idea: that human beings strive toward the realization of their potential. Goldstein (6) in physiological, Rogers (14) in personality and therapy, Maslow (9) in motivation, and others, have used such terms as *self-actualization* and *self-realization* to express this concept. In an important recent article, White (16) has summarized the research and writing from such diverse areas as animal experimentation and psychoanalysis as a basis for suggesting the concept of a basic human drive toward competence. This idea, in conjunction with greater emphasis on process as argued by Allport (1) and with some of the research on perception such as the Ames (3) demonstrations, has tremendous implications for learning.

The assumptions to be presented attempt to take into account the recent findings. The paper will not proceed logically from the assumptions to the following concepts, however. The concepts developed are consistent with these three assumptions: man is purposeful, man is becoming, and man is aware-ing.

Man is purposeful

Fromm (5) argues that a distinguishing feature of man is that, unlike all other

animals, man must build his relationship with the world in which he lives. The characteristic and instinctual drives of other animals have evolved in direct integration with their environment so that their survival activities—food gathering, sheltering, mating—are transmitted through the chromosomes. Animals other than man are part of Nature —they cannot do an "unnatural thing." Man in contrast is born helpless and with few behavior patterns built in for coping with the world. He is born into a cocoon of culture which protects him from meeting "nature in the raw." Man must learn to cope. Over the generations before him, complex patterns for meeting his physiological requirements and for protecting himself from danger have been developed. But they are imbedded in his culture, not in his organism. They have been transmitted independently of his biological inheritance. These patterns of culture form the general outline of the relation he will build with Nature, but each individual must internalize and develop his own unique pattern from this model. Unless he does so he cannot survive in Nature. This fact lays a clear foundation for the assumption that man *must* strive to establish an adequate relationship with his environment. This purpose gives direction to man's life.

Man is becoming

When we try to describe the functioning of human beings we find that our language betrays us. We have too many "thing" words and an inadequate supply of words or phrases to describe on-going processes. Thus, we have the word "consciousness" to refer to what is really a flowing continuous process which can never be frozen into an entity. There is a tendency to speak of a process as though it were an object and to overlook its dynamic qualities. It is difficult to define something which is continually in motion, and it is only natural to try to freeze it for a moment so that we can see it more clearly. The danger lies in the fact that frozen processes are often markedly different from on-going processes and so the real nature of them eludes us in still life.

In attempting to understand human beings this difficulty in describing processes shows itself most clearly when we discuss changes and the direction which changes are taking. It is difficult to make phrases which communicate about the "ordinary" or the "normal" state of an individual without obliterating the idea that everchangingness is the ordinary and normal. It is difficult to speak of an individual as having an identity without losing the realization that the "identity" is always changing into something else.

When the individual is functioning smoothly, when it is making the necessary changes to maintain itself and become more effective, only then does it appear to be an identity to an outside observer. Only as it undergoes necessary changes does an individual have stability of process so that it can be recognized as an identity. The very changing maintains the identity of the individual which gives the observer the feeling of stability about it. When one talks about an individual changing, then, he must be referring to the altering of its process of change. It is only

when there is some disruption of its integrated changing that there appears to be something "not ordinary" or "not normal" about an individual.

For example, when the normal metabolic processes are taking place in a body, the individual appears "the same," though the cells making up the body are 98% changed in the course of a year. When something interferes with the metabolic process, however—when some change does not occur as expected, when something about the process of change changes—then the individual is different and may even lose its identity.

An example in more behavioral terms is seen when a child in school first faces the situation of functioning with thirty other children in a classroom. In order to continue to appear as a happy, well-adjusted child, he must develop patterns of relating which are changes in him. If, however, the changes in him are a smooth and integrated process, he will continue to appear to observers as "the same" happy, well-adjusted child.

The unchanging thing is dead. A person must change continually just to stay as he is!

Another aspect of the everchangingness of life is a matter of relatedness. Even if an individual could stay "the same," its environment would change over time so that the individual would have different relationships, and would in its functioning be a "different" person.

It is time, therefore, to abandon an assumption common to the stimulus-response theories which have so strongly influenced education. The assumption is that the human organism is a passive object shaped by the impact of more or less random stimuli, as if the organism continues to respond as it has been shaped until, by chance, some new stimulus reshapes it. The assumption made in this paper does not allow humans to be seen as the *inert* product of past casual factors. Descriptions of behavior must show the human as purposely and continually selecting stimuli to which to respond so that humanness is an integrated process of directed change.

This complex process of continuous changing-into, this concept of man as being-from-past-through-now-to-what-will-come-of-it is what Allport and others have described as man's *becoming.* Man must become ever more adequate in relationship to his environment.

Man is aware-ing

The key to man's purpose and becoming is the process of his being aware. This concept is re-emerging in experimental psychology and is gradually being considered as essential to our understanding of learning. As Mowrer points out, "—it is perhaps not too much to say that we have now reached a point at which, if consciousness were not itself experienced, we would have to invent some such equivalent construct to take its place" (10).

The lack of words to denote process is felt keenly at this point. It is not correct to say that man *is aware;* rather he is continually *aware-ing.* Man is not just conscious of his environment—he is continually relating through his behavior to what is happening to him in

his environment. He does not just know about the environment—he is continually rebuilding, extending, and applying his knowledge through the process of aware-ing.

Aware-ing is an innate potential of man. It must go on, just as the heartbeat or any autonomic process must go on. It is a process essential to life. In a mechanical sense, humans are equipped with elaborate feedback devices which meter the organism's state of becoming, internal and external. The human is *knowing,* and only in knowing does he continue to survive. It is the aware-ing of alternative courses of action, of consequences, of states of one's own organism and the environment which shapes behavior and learning.

All that each of us is knowing integrates with what we have been knowing. We cannot have a new knowledge except as what we know tells us what the new knowledge is. This complicated idea has been expressed by several poets. For instance, Gibran has it:

> *No man can reveal to you aught but that which already lies half asleep in the dawning of your knowledge (18).*

Each man constructs his knowledge, but this is an integrative, not an additive, process. In each new experience there must be elements already in the aware-ing process if the knower is to integrate the newness.

Thus each of us forms his own unique perceptual field. Each of us from birth recognizes that relating to certain aspects of the environment has more consequences than relating to others. Some aspects of what we encounter are more involving. Some parts of the world we live in are more important to us and take on different values. These values are integral to our aware-ing. The world each perceives is a world of private meanings real only to each individual. Aware-ing binds every man to the world he lives in, but not to the same world known to any other man.

It is assumed that there is a world to be known. The generalizations of the physical sciences and those which we strive for in the social sciences can be valid only if there is something about which to generalize. The fact that it is possible to make such generalizations leads also to much overlap and similarity among the perceptual fields of different individuals. This does not, however, change the fact that each individual is uniquely aware-ing of even these common generalizations while he also includes in his perceptual field many ideas which may not be shared by others.

Much studying and theorizing about learning has taken place in which the organism and the environment with which it interacts are separated. Mowrer (10) argues for the position of studying learning separately from the environment by making an analogy to genetics which studies the mechanisms of heredity without concern for the environment in which the individual matures from his hereditary base. The position is certainly a plausible one; yet it is one which seems to lead to theories which cannot be applied to live situations, and still leaves the major problem of the organism-environ-

ment relationship untouched. As Sears points out, this kind of description ". . . is systematically acceptable only if other variables are added that will, together with internal personal properties, specify what kind of behavior can be expected for him under some specific circumstances" (15). A wise course would seem to be the inclusion of "relationship" as a key concept in learning from the beginning.

In the following development an attempt is made to do this by including the world within the perceptual field of the individual. From this point of view the learning process is continually one in which the individual is guided by the relationships between the drive of the organism to express itself and the context within which this expression must take place.

Thus there are two aspects of aware-ing emphasized in developing a theory about how man learns. The world each man is aware-ing about is one unique to him—he is the only one who knows the world just that way. And each man must be aware-ing of the world in some way. No matter how distorted his aware-ing may be, a man must try to relate to the world known to him or die.

Expression of the Self-Concept

Consistent with the three assumptions which have been described, there are several concepts of how man's expression of self leads to learning. These concepts have to do with (1) how each individual develops a perceived-self-in-the-world and a concept of adequate-self, (2) how a lack of congruency between these self-concepts is the basic source of motivational states, (3) how satisfactions are defined which will express motivational states and create new relationships in the self-systems, and (4) how, given a particular process potential, each individual expresses his adequacy in action.

Two self-concepts

There are already a number of researches which point to the usefulness of the self-concept. Rogers (13) in various studies of the progress of therapy used a Q-sort to measure discrepancies between the ideal and perceived selves. Bills (2), with a different instrument of his own, showed that self-acceptance is related to a number of other variables important in good human relationships. Orzeck, McGuire, and Longenecker (12) in an ingenious study using hypnosis, demonstrated that there are changes in self-concept associated with induced moods of elation and depression. Hebb, in his Presidential Address to the APA, pointed out that evidence is accumulating that there is "a pure fantasy of an immaterial self . . . with effects on behavior" (7). Recently, Wylie summarized the research on the self-concept which has been conducted in the past two decades. She stressed the many inadequacies in the theorizing and the research designs which have characterized attempts to use self-concept theory. Self-concept theories attempt to deal with variables which are not presently included in general explanations of behavior. With few exceptions, such theories have not yet had notable success, but Wylie indicated ways in which work in this area can be greatly im-

proved. Thus, it is felt that there is an empirical basis for the concepts discussed in this paper.

As a direct consequence of aware-ing, an infant begins to recognize some relations as self and some as not-self. This recognition expands until there is a fully developed concept of a self-in-the-world which is a chooser, director, actor, knower—as an experiencer of consequences.

By the time a child comes to school he is aware-ing in a rudimentary fashion, a kind of two-fold picture of himself. From his experiences with himself and the way other people and things respond to him, he has developed notions about the kind of person he is, including the way he looks, whether or not he is lovable, and what kinds of things he can do. Included in the definitions of what he can do are descriptions of the things in the world he can use to do what he does. These notions are the perceived-self-in-the-world. How the "world" affects this self is *meaning.* At the same time he has developed a concept of what he should be. The same experiences with the world have led him to develop a concept of an adequate self. Rightly or wrongly, he thinks that if he could be like this concept of an adequate self he could handle all the problems of life!

The perceived-self-in-the-world and the concept of adequate-self are unique for each individual. It is paradoxical that the concept of self arises as each person recognizes his uniqueness, but each person finds it difficult to realize that each other person is entirely unique, too, and lives in his own perceived-world.

For any individual neither the perceived-self nor the concept of adequacy is entirely "real." Each person has lacked some experiences for learning, and each person has distorted some of what he "perceived" in order to integrate the perceptions with his concepts of self. Probably it is easier to distort the perception than to change the self concept in many cases. Also, in the process of differentiating and categorizing relationships which develop, some "meanings" develop which are part of the self-concepts but are not integrated into the system. That is, certain "feelings," certain unsymbolized "meanings," develop at a time when there is nothing else in the self-system with which to relate them. Many feelings about *mother* and *father* have this un-named-meaning aspect. Thus, with some lack of learning, some distorted perceptions, and some unintegrated relationships, the perceived-self-in-the-world and the concept of adequacy both include "unrealities."

The theoretical presentation up to this point, with its emphasis on aware-ing, has been concerned primarily with conscious processes. These are the processes which, it is felt, enable us to understand learning. When one moves broadly into the field of behavior it is clear that factors other than conscious processes also make a contribution. These are important in understanding learning because they contribute something "extra" to the outcomes of attempts to foster learning. Sometimes this "extra" may be a distortion of the learning and upon occasion it may even serve as a block to effective learning.

The source of this extra contribution to behavior seems to lie in the area of feelings. Accompanying all conscious processes of interaction with the environment there is some organismic reaction to internal conditions in the organism and a reaction to the consequences which the individual's actions cause to be fed back into awareness. It is believed that these feelings are the basis of all "meanings" which an individual develops over time. Starting at birth, or before, changes in the internal or external environment give rise to feelings which are inherently pleasant or unpleasant. As the individual is able to bring into awareness the relationship between the causal condition and the feeling, the relation can be symbolized and dealt with consciously. To the extent that this relationship cannot be awared and symbolized it cannot be dealt with consciously. Nevertheless, the feeling, when present, contributes to the context in which action must take place. Just as a rough spot in the road affects the course of the vehicle even if it has not been seen, so a "rough" feeling affects the course of an individual's actions even though not awared. If the driver sees the rough spot he can take action to avoid it. If an individual becomes aware of a feeling and can symbolize it, he can plan a course of action which includes the meaning which the feeling has for him.

In other cases, when recognition of the feeling is in conflict with important elements of the individual's concept of adequacy or of his perceived self, individual therapy may be necessary before further learning can take place.

A point to be stressed is that the concepts of self are processes which must be maintained and which will change over time as new experiences are absorbed into the system. That is, the selves are becoming continually, but no change can occur which does not grow out of the self in some way.

Motivational states

The motivational state, as we are using the term, is what has been called a need. The term *need* has come to include so many meanings that it has almost lost meaning. Since it is desired here to employ the concept in a particular way, and exclude other meanings, the focussing label *motivational state* is used. A motivational state is the necessity to make congruent some discrepancy in the self-system.

A motivational state develops whenever there is a discrepancy between units in the self-system. Discrepancies may develop in a number of ways in a system which is not functioning to full potential, but the most direct and usual lack of harmony is between the perceived-self-in-the-world and the concept of adequate-self.

To some extent the perceived-self-in-the-world and the concept of adequate-self overlap. At any time in the process of becoming, part of the perceived-self is an adequate-self. However, the two selves are usually discrepant in many ways. Basically it is the drive to become the adequate-self, to remove the discrepancies between what is perceived to be and what should be, that we call a motivational state.

Whenever there is a gap between what a person perceives himself to be

and what he believes to be adequate, he must try to close the gap. Therefore, a person is motivated to undertake any development of which he perceives himself capable if it will enable him to perceive himself as more nearly congruent with his concept of adequacy. (What is developed later as a theory of learning does not exclude unconscious motivational states. Therapy is the process by which the unintegrated "feelings" and "meanings" are brought into awareness so they can lead to adequacy.)

This way of looking at motivational states puts a special emphasis on the expression of self rather than on organic coping with the environment. To be sure, man makes a great variety of adjustments, he uses his world in complex ways, but all the while these coping techniques are a means to the expression of his values about adequacy.

For example, if a child who does not know how to read sees his adequate-self as one who can read, then he is in a motivational state in which he demands of himself that he try to read. If his perceived-self is that he reads as well as an adequate self needs to, then he will read but he will not be motivated to improve his reading. Or, if his notion of adequate-self precludes acting like a baby, he will be motivated to avoid reading baby terms such as "bow-wow-wow" or "quack-quack-quack." In whatever way the child relates to the task of reading, his efforts are an expression of his motivation to make his perceived-self congruent with his adequate-self. In the same way, it can be shown that a man chooses a mate, goes to college, steals a car, or engages in any behavior because he is motivated to remove a discrepancy between what he perceives himself to be-in-the-world and what he believes he should be.

Definitions of satisfaction

Before an infant develops a concept of self his motivational states arise from lack in the supply of the basic conditions of life. To maintain process and to grow he must have such things as food, water, oxygen, warmth, safety and perhaps fondling. But as he turns his potential for aware-ing toward purposeful becoming, the child begins to recognize a variety of satisfactions in the basic supplies. He does not identify with all of them equally; for example, he finds some more or less complete, more or less enduring, more or less consistent. Soon all the basic supplies are overlaid with values and valences. Such a seemingly simple supply as food becomes wrapped in definitions of what is acceptable to eat, when, where, and how. It is not enough to be nourished; one must have the kind of food which meets one's definition of satisfaction.

Of course, the definitions of satisfaction soon become extremely complex and individualized. Supplying the essentials for staying alive becomes a minor part of motivational states. For every kind of discrepancy between the perceived-self and the adequate-self, a child develops a definition of what it would take to close the gap. Such definitions of satisfaction are sometimes called goals. In process terms

goal has the objectionable connotation of *end.* Perhaps *route* or *path* more nearly conveys the idea. To illustrate, an individual may see the learning of some skill, the taking on of some knowledge, the gaining of a certain job, the performance of some task, or the acquisition of some thing as what it would take to become his adequate self. Two individuals with quite different motivational states might realistically undertake the same-appearing route to satisfaction. On the other hand what makes a definition of satisfaction for one person may make no sense at all to another. Thus, a boy may steal a car because he is motivated to prove his masculinity, while the very act appears as irresponsible and immature to others.

Again it is emphasized that the definitions of satisfaction are in terms of self-expression, they are the creative direction to one's becoming; they are not simply coping techniques, though they do accomplish the adjustment of the individual and his world. When deeply related to the self-systems as expressions of long-term motivational states, definitions of satisfaction become seemingly fixed patterns of behavior. Thus a man may choose a mate and remain married to her through all his mature life as a continuing expression of how to perceive himself as adequate. However, the definitions of satisfaction are also in process, so they grow and change. Any alteration of the adequate-self must be accompanied by new definitions of satisfaction. The definitions may appear to be unrelated to previous ones, but they cannot be unrelated to previous notions of self.

Process potential

The descriptions presented thus far have pictured a chain of relations that result in behavior. Individual organisms, within which these processes are presumably taking place, vary widely in certain biological characteristics which condition and limit psychological development. For an individual, his particular combination of intelligence, state of health, stage of maturity, degree of strength, state of fatigue, and many other such qualities will be called his process potential. The self-system lies within a process potential which shapes it in various ways. Sometimes the process potential "opens" the world of experience; for instance, a person of high intelligence can make much use of what happens. On the other hand, process potential sometimes is a limiting factor to experience; for example, a blind person cannot have certain experiences.

Choice of behavior

For any definition of satisfaction there is some universe of behaviors which, given one's potential, will move one along a route to self-expression. However, not all of these possibilities are open to an individual. Only some have been made known by one's culture, only some are made available in one's social group. Of these, only some are part of the individual's perceived world. And of the behaviors recognized by the individual as meeting his definition of satisfaction, only two or three may be appropriate to the current situation. Between these two or three there are weightings or valences which deter-

mine a specific course of action. Part of this weighting is an anticipation of the consequences. This is essentially the same construct as Sears' *expectancy* (15) or Hull's *anticipatory goal response* (8).

Aware-ing

Once an individual has made a choice he acts in a way which is appropriate to this choice. This action has consequences and these may take several forms. He may be able to observe his own act directly through seeing it, hearing it, or feeling it. Perhaps more frequently, the consequence which is experienced is the reaction of some other person with whom he is interacting. This outsider, who has a self-system of his own, evaluates the action in terms of his private definition of satisfaction and responds to the first person. This response may indicate approval or disapproval, agreement or disagreement, or in some other way provide data to the first person. This experienced consequence is fed back into the self-system and is awarded or evaluated. The essential criteria for evaluating the consequence come from the two self concepts.

First, the consequence is compared to some aspect of the perceived self. If it is to be accepted for further evaluation it must be relatively consistent or congruent with this percept. For example, the student in school who has been receiving "C" grades will find any new grade in the "B-C-D" range as believable or approximately congruent. He can then check this datum against his concept of adequacy to see if he has been successful or unsuccessful. However, if the grade is an "A" or an "F" it may be too inconsistent with his concept of self and it cannot be directly accepted as data. The discrepancy gives rise to feelings to which he responds by trying to distort or interpret the data in some acceptable form. He may characterize the grade as being a lucky chance or as being unfair and, therefore, it cannot be used as data to test adequacy. The consequence will thus be changed into a form which is consistent with self but not used as data to test adequacy. In the event that the consequence is too discrepant with perceived self it may be blocked completely. If a more extreme response is received in the example above, one in which the teacher appended a note saying that the student's work was completely unacceptable and that he doesn't even belong in school, this would give rise to such strong emotion that the data would be lost. The emotion would be vented through some attack on the teacher or on the school but the actual feedback of the data would not be symbolized and would be, therefore, denied to consciousness. A person who made such an interpretation of the consequences might learn something different relating to self and the school but he would be blocked from further coping or learning along the original dimension which the act was attempting to achieve.

In the event that the consequences are not incongruent with perceived self they are compared to the criteria derived from some aspect of the concept of adequacy. If this comparison shows that progress has been made or that

the act was adequate the individual will have pleasant feelings. He will perceive himself as more adequate and will be motivated to continue choosing and acting in a similar manner. Parenthetically it may be said that this is not a hedonistic point of view. Pain may be part of the experienced feedback, but if the interpretation is that the individual is becoming more adequate it may still result in pleasant feelings.

Model of expression of self

A diagram may help in understanding the process which has been described. Figure 1 is a simplified paradigm of the process of becoming. This is a perceived self motivated to become an adequate self.

In this diagram, the curved shapes represent aspects of the process which are internal to the self-systems. The rectangular shapes are perceived as "not-self" but are integral to the process of becoming.

At the left of the drawing, couched in the process potential, the two circles which are slightly overlapped designate the two self systems. Part of the perceived self recognizes its adequacy, but the greater parts of the two selves are discrepant. The discrepancies set up motivational states. Something in the system must change so that the two selves may become more similar. A discrepancy between the two selves also

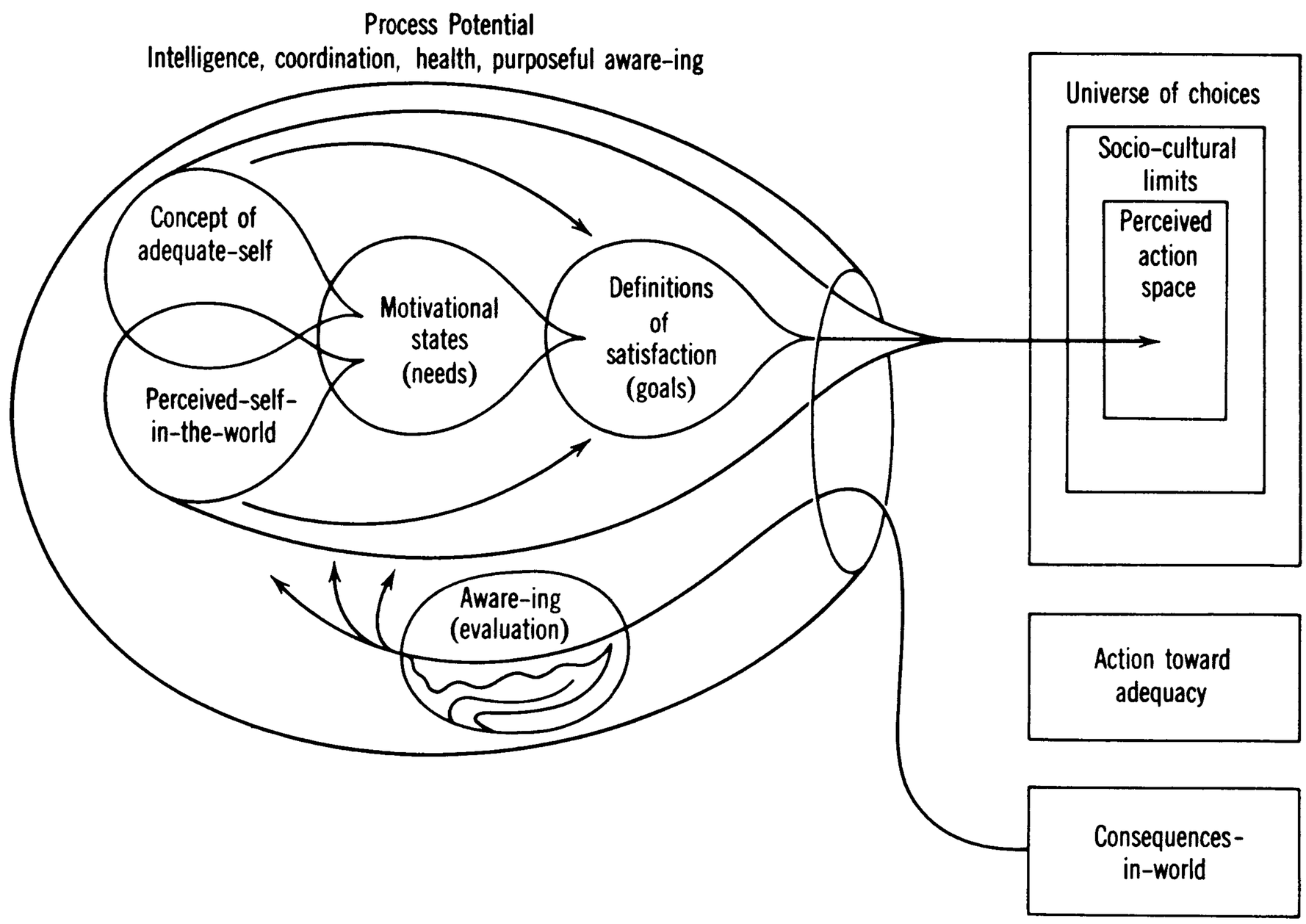


FIGURE 1. A model of the expression of self. When there is a change in the existing relationships between the self and other elements in the process, there is learning.

determines the definitions of satisfaction that are selected. However, it is the motivational state which determines the degree of force with which the satisfactions will be sought. Next, process potential supplies a final sharpened focus for the relationship of the self-system to the environment. While the process potential is largely inherited, and/or organic, and autonomic, it may change with experience and is subsumed in the overall process. It cannot function except as a part of becoming.

At the right of the diagram, the top set of rectangles represents the behaviors possible to the system. For any individual the perceived action space is relatively limited in comparison to all of the actions that are "really" open to him. In the perceived action space one course of action is determined which meets the situation and the definitions of what will satisfy the motivation to make the perceived self the adequate self. This action becomes part of the system. And then, as part of the action in a situation, there are consequences.

Some perception of these consequences feeds back into the system where it is awared. The awareing process consists of comparing the feedback data with existing aspects of the perceived and adequate selves. If there is a discrepancy between these data and the concept of perceived self the data will be distorted or blocked and emotion will be induced. The emotion will be handled by appropriate defense mechanisms. If the data are congruent with self they will be compared to criteria of adequacy and the resulting evaluation will begin the cycle over again.

Learning

Now it can be shown how learning fits into the process that has been described. Learning is part of the process, but it pervades the total process and cannot be drawn at any one point in the picture. Learning has occurred whenever there is a change in any relation among the parts of the process of self-expression.

Human learning is the changing of relations between a self and its perceived-world as the self is expressed in striving to become adequate. Two kinds of learning which fit this definition are distinguished and labeled *instrumental* and *intrinsic.* The significance of intrinsic learning is in the purpose, direction, the "why," of behavior. The significance of instrumental learning is in the content, the procedure, the "how," of behavior.

This terminology parallels the distinction made by Olds (11) between instrumental and intrinsic reward-values. As he pictures the process, his "growth and decline of reward-value" is learning, and he describes two kinds as follows:

There is an important distinction to be made between two kinds of reward value that an object may have. The one kind is symbolic, instrumental, extrinsic; the other is real, ultimate, intrinsic. The one is that which it has as a token or instrumentality exchangeable for some more ultimate reward; the other

is that which it has as an ultimate reward in its own right . . . The kinds of operations which change intrinsic value may produce small . . . changes in any given experiment, but they may cumulate over the life of the organism producing, eventually, basic changes in its make-up. On the other hand, operations which change instrumental value, although they may be easy to apply and although they may make big changes in short-term experiments, may produce results which are so ephemeral in the long run as to be unimportant influences on the basic make-up of the organism (11, p. 28).

As the expression of self takes new forms, using new skills, or new means for accomplishing the individual's purposes, instrumental learning has occurred. The self-system maintains the same adjustment to the environment, but *how* this is done has changed. The individual uses different instrumentation. On the other hand, as there are changes in the perceived and adequate selves, the individual takes on new "ends," and intrinsic learning has occurred. What the individual is striving to become has changed. This gives rise to new motivational states and definitions of satisfaction. Intrinsic learning necessitates new instrumental learning, but new instrumental learning does not necessitate new intrinsic learning.

In the paradigm of self-expression, if the changed relationship is among the rectangles, the learning is instrumental. The learner may find new facts which open more choices of action to him. He may be able to operate faster, with more efficiency, through shortcuts, or with new tools. He may put data in new order. He may use data so there are more consequences. He may learn of consequences about which he has been unaware. He may increase his skill in evaluating. As long as these changes occur with no effect on the self-concept, or with only slight reinforcement of the self-concept, then the same self is behaving with improved instrumentation. The becoming of the learner is more efficient, the process is facilitated, but the direction of becoming is not affected.

Changes in the other shapes in the diagram, however, involve intrinsic learning. Intrinsic learning means that parts of the self-system are in a new relationship. The learner perceives himself as a new self. He realizes that some new element must be included in adequacy. He recognizes that satisfactions must be obtained by different routes. Any change in this part of the process changes the self-system that is expressing adequacy, and becoming must take a new direction.

Examples of instrumental and intrinsic learning

In order to show how this definition of learning and the distinction between the two types of learning clarify what goes on in the school setting, several examples are given. Let us consider some of the great variety of concepts of perceived-self and adequate-self giving rise to instructional states and some of the different definitions of satisfaction when children learn to read.

Because of past experiences at home and in school a child comes to have a concept of the adequate self as being one who can read. His present concept of himself is that he cannot read but that he is capable of learning to read. This discrepancy between the concepts motivates him to learn to read and he defines satisfaction as the ability to pick up a book and read it. With the help of a teacher he relates printed symbols and his perceived-self, and relates symbols to each other. He feels the satisfaction of being able to recognize some words, of getting indications of success from the teacher, and possibly a warm response at home. In this case he will continue to be guided by the teacher and will gradually change his self concept to see himself as a reader. At length, as this concept more nearly coincides with his concept of adequacy as a reader, the need to learn to read will disappear. This may occur at the same point at which the school feels he is an adequate reader and everyone will be happy. If his ability is below the school goal then he must have experiences which change his concept of adequate self before he will again be motivated to learn more about reading. All of the concepts discussed in the previous sections and illustrated in Figure 1 are included in this example.

Now let us look at two variations of this example. If the child comes to school from a non-reading home his concept of an adequate self may not include the ability to read. There will then be no motivation to learn reading. The child will appear resistant and distractable when the teacher makes his suggestions. The child must have supportive experiences which will build up the concept of an adequate self as a reader. Probably the word "supportive" is the key one in this last sentence. If the child has experiences in school which build a concept in him that he is worthy and important then he will respond to his awareness that others around him are learning to read and that school sees reading as important, by taking on the concept that an adequate self is one who can read.

A second variation of the example may be that the child has the concept of reader as a part of his adequate self and proceeds to do as the teacher suggests. As a consequence of this, for whatever reason, the teacher tells him he failed and that he hasn't listened or worked hard enough. Then the experience will not be a satisfying one. He did what the teacher suggested and failed; the teacher has not suggested any alternative actions which satisfy him. He may then decide that he is a self who can't read. Several experiences like this will confirm this view of himself and he is blocked from becoming an adequate-self. This will, of course, continue to fester and lead to further distortions of self and possibly to the ultimate rejection of reading as necessary to an adequate self. Then, all motivation will be gone for learning to read.

In these examples we have oversimplified for the sake of clarity. Actually, an individual will have discrepancies between perceived-self and adequate-self in many areas, and his motivational

state is a complex one. Several definitions of satisfaction will influence the action which an individual takes in any situation. Thus, in the example, the child is motivated because he sees reading as part of an adequate-self. There are other children who have not yet seen reading as essential to an adequate-self yet who might try as hard to read because their concepts of adequacy require a teacher's approval. That is, an adequate-self is one who is approved by authority figures. The child does not perceive himself as thoroughly approved and so he is motivated to please the teacher. Such a child will follow the teacher's suggestions that he should do things which will make a reader of him. Learning how to read may be only incidental at this point. For many children both of these motivations are operating as well as motives to be like their classmates and to be approved at home. Each of these motivations may reinforce the others.

In the illustrations above both intrinsic and instrumental learnings occur. In the first case the child learns the instrumental skill of reading and also makes the intrinsic change in self-concept from non-reader to self-as-reader. In the case of the child from the non-reading home, the instrumental reading skills will not be learned until the intrinsic learning of the adequate-self-as-reader takes place. In the case of the child who failed, the instrumental learning did not occur, but two intrinsic learnings could result. First he might learn that he was not an able self. Living with such a concept is uncomfortable and might lead to a concept of adequate-self which no longer required that he could read.

Such illustrations point up the fact that aware-ing of one's experiences brings three kinds of evaluations into the self-esteem (1). The evaluations may confirm the various perceptions which are already part of the self-system (2). Some element in the evaluation conflicts with either one or both of the selves. The new element cannot be integrated into the system and must be rejected (or distorted) in order to preserve the system (3). Evaluation enables the individual to see himself differently or in some new way which is not in conflict with any existing percepts. The "entirely new" element, of course, is a newness so related to the system that it can be known and integrated.

Of the three kinds of evaluation described above, only the last two result in intrinsic learning. In the last kind, the new element is integrated directly into the system. The second kind of evaluation results in intrinsic learning only after the conflict is handled in some manner.

In the discussion up to this point intrinsic learning has been emphasized. Instrumental learning is also important. The schools are legitimately concerned with instrumental learning and must devote a major portion of their time to fostering it. The laboratory experimentalists such as Hull, Skinner, Spence, Guthrie, Estes, Mowrer, and others, with all their differences, have had great success in isolating the variables involved in instrumental learning. The concepts of conditioning, reinforce-

ment, secondary reinforcement, and contiguity, contribute to an understanding of how the instrumental actions of an individual are modified. Schools must be concerned that students are increasing skills, memorizing facts or systems, and developing almost automatic responses to certain situations. Principles from the laboratory experiments explain how the process potential is shaped into instrumental learning. Instrumental learning will not develop, however, if the role of intrinsic learning is ignored. In the past, children have been finding the intrinsic significance for themselves in the learning experiences provided in schools. Whether or not they learned what was "taught" has depended on their finding a use for the "learning" in expressing their selves. As long as intrinsic learning has been ignored and its course determined inadvertently, teachers have been baffled by children who have not learned as they were expected to.

Teacher–student relations

The teacher as a behaver can be represented by another diagram similar to Figure 1. He too has a concept of adequacy and perceived-self and all of the things which follow from this. Part of his concept of adequacy contains values with regard to being a teacher and he will perceive of himself in certain ways as a teacher. These conceptions will determine the definitions of satisfaction which he formulates and will guide the choices of behavior which he makes in the classroom. It is at the point where the teacher takes some action as a result of his choices that he affects the learning of the student.

For the student, the teacher exists as many relationships—adult, authority, status figure—all in interaction. Through his various experiences with adults who have authority over him, the student's concept of adequacy will include some recognition of living up to what is expected of him by authorities. As the teacher tells a student to be quiet, or that he has done good work, the student perceives these as consequences of his relationship with the teacher, and these consequences affect his evaluation of how well he is achieving his end of becoming an adequate-self.

It seems clear that the teacher plays a major role in fostering instrumental learning. The choices and actions which he encourages a student to take will increase skills and knowledge. It is less clear how a teacher can foster intrinsic yearnings, and most teachers pay little attention to this aspect of learning. If the child is learning instrumental behaviors most teachers are content. If a child is not learning instrumental behaviors, teachers see him as a learning problem to be referred. Yet, it seems clear that teachers do affect intrinsic learnings and that some of these learnings foster further instrumental learnings while others tend to impair or block them.

We have given examples of intrinsic learning in school with the reading examples above. There could be many more. Some students learn how to relate to teachers in such a way that the adequate-self is one who can get by

without doing any work, or getting caught. Other students come to see the adequate-self as a person who is like the teacher. Some perceive themselves as stupid or incapable of a skill even though their process potentials do not limit them in any such way. This is surely the problem of some underachievers in school. All of these examples illustrate intrinsic learnings which have probably been fostered, inadvertently, by teachers. A major implication from this whole theory of learning is that teachers must become aware of and understand students in terms of intrinsic learning as well as instrumental learning.

It is evident that the factors involved in intrinsic learning are very complex since they involve the total personality. Teachers do not have the time, nor in most cases the training, to determine what each child's concepts of perceived-self and adequate-self are. How then can a teacher take intrinsic learning into account systematically? Fortunately, this can be done without detailed knowledge of the individual child.

The theory being presented holds that the actions which an individual takes are an attempt on his part to reduce the discrepancy between his concepts of an adequate and perceived selves by becoming an adequate-self. Therefore, his actions can tell us what the discrepancy is that he is working on. By his actions he is communicating to us if we can only hear what he is saying. If we do hear accurately and respond accurately we will be helpful to him.

An example of good communication may help to clarify what is meant by responding accurately. The following incident was reported by a teacher in an anecdotal record about a girl in her fifth grade class.

Sept. 17—The class was organized into three reading groups today. Becky was put into the middle group. When the books were given out, the children were looking at them—getting acquainted with them. Becky came up to me and said, "This is a fifth grade book, isn't it?" I replied, "Yes, it is, Becky. Doesn't it look interesting? All the stories have been selected to appeal to fifth grade boys and girls." She relaxed as I talked and said, "I didn't want any fourth grade book." She returned to her seat smiling and hugging the book.

Even without the inflection of her voice and other clues which the teacher had, it seems clear that Becky was communicating two things. At the content level Becky was asking a question about the book. At the feeling level she was indicating some anxiety. The form of her question, "this is, isn't it?" was saying, "I need some kind of reassurance." The teacher responded to both of these messages quite accurately by reassuring her that it was a fifth grade book. Becky's comment and actions as she went back to her seat confirms the fact that the teacher had communicated.

Communication is basic in relating between children and teachers. Communication means that there is enough sharing of experiences so that many symbols mean somewhat the same to the two people in interaction. Commun-

ication makes it possible for the behavior which brings satisfaction to the teacher's motivational states to be the same action that meets the child's definitions of satisfaction. What makes the teacher adequate must also facilitate the child's feeling adequate.

Of course, teaching requires that the teacher have command of a great deal of information. What he has learned should be put in the student's service. However, what the teacher supplies to the learning situation must emerge on demand of the student's motivational states if it is to facilitate the student's learning. If it emerges, instead, by edict of the teacher's sense of structure, then the teacher will fail to communicate with most students. Most of what the teacher contributes to the learning situation is not of his knowledge, but his self. He is teaching only if he can relate to student's self-esteems so as to facilitate student becoming. He is teaching only as his self relates to students without restrictions, limits, or conditions on their motivational states. He must be becoming, without blocking the becoming of students.

Most intrinsic learning occurs in a relationship with another human being. A child's self-esteem must maintain a balance in relation to a number of significant adults. It is important that the adults be able to support the child's efforts to become adequate without the adult's approval becoming a requisite to the child's concept of adequacy. If the approval of the teacher becomes a part of the student's concept of adequate-self-in-school—and it certainly does for many students—then the student will learn whatever is demanded by the teacher if thereby he perceives himself as approved. If the teacher then withdraws, restricts, or qualifies approval, it creates a discrepancy between the student's perceived-self and concept of adequate-self. He must close such a gap, but the learning resulting from his effort to do so is not intrinsic learning, but instrumental. His behavior enables him to continue adequate as a teacher-pleasing person. He can forget the behavior when the particular teacher no longer affects his adequacy.

A teacher should be able to disapprove of a student's choice of behavior or even his definitions of satisfaction without the disapproval constituting anything more than information for him to consider. He may come out with many intrinsic learnings through which he becomes more adequate in relation to his world through the help of a particular teacher, but the mere learning of how to gain her approval is relatively unimportant to his ultimate adequacy. It may even block longer-range learnings as the student limits his actions to teacher-pleasing behaviors.

In terms of this theory a teacher's concept of an adequate self should include at least two dimensions. One of these is that an adequate teacher is understanding of children. The other is that an adequate teacher is a facilitator of learning. These two are actually linked. The understanding teacher is continually trying to infer, from the student's behavior, in what way the student is striving to become more adequate so that the teacher can provide the response (data) which will facilitate the student's learning.

Another implication of this theory is

that learnings, instrumental or significant, are not available to a student just because the school, teacher, or text says they are. The student can learn only as experience is related to his motivational states and definitions of satisfaction. There are some things a student cannot learn until he has a different self.

While the experiences that a child has in school will lead to many changes in his concepts of an adequate-self and his perceived-self, these changes will occur only as his evaluations of experience do not conflict with existing elements in the self-esteem. It is possible that some of the conflicts can be resolved within a supportive classroom atmosphere, but many conflicts probably cannot be. Where this is the case *learning cannot occur.* Then only through successful therapy can the conflict be resolved so that learning can continue.

Society does set purposes for schools, and society does demand certain products from schools. This does not alter the privacy of a student's definitions of satisfaction. Only as the demands made on him become part of his private perceptual field, and only as the purposes for schools are interpreted by his personal adequate-self-in-school, is his learning affected. Society can transmit a culture only as unique self-systems can use it.

Implications for Research

Two different kinds of research are suggested by the foregoing theory of learning. In the first place, as Olds (11) points out, a new theory requires a number of studies to test its various definitions and to make more operational the concepts it introduces.

In the earlier discussion rough definitions have been presented in which intrinsic learning is concerned with the direction and strength of behavior and instrumental learning with the content of behavior. Direction, strength, and content of behavior are directly observable variables. Most of the self-concept studies (17) which have tried to measure the intrinsic factors of direction and strength have employed instruments administered only once. Clearly, direction and strength of behavior can be seen over large time segments, and it may be that, at first, two or more measurements over time will be necessary to identify these intrinsic variables. When results from studies which measure intrinsic behaviors over time are related to the more directed descriptions of adequacy and perceived-self, it should be possible to determine intrinsic variables in a single measudement.

A second area of research relates to the implications of this learning theory to school practices. Following are hypotheses which, if tested, might bring important educational developments:

(a) If the structure of content in a course emerges as uniquely related to the motivational states of each student, the direction, strength, content, and persistence of student behavior will change more than in a course the content of which has been pre-structured without specific regard to the individual students.

(b) Where a teacher's need for adequacy leads to classroom behavior unrelated to the students' definitions of

satisfaction, the direction, strength, content and persistence of student behavior will not change as much as in other settings.

(c) Becoming adequate, if defined by students as pleasing the teacher, will result in learning of an ephemeral character; i.e., the changes in direction, strength, and content of student behavior will disappear when the given teacher is not a factor in the situation.

(d) Becoming adequate, if defined in terms of student satisfaction, will result in learning of lasting character, i.e., the changes in direction, strength, and content of student behavior will persist through time and in a variety of settings.

(e) Teachers sensitized to the concepts of adequacy and the perceived-selves being expressed in the behaviors of students will facilitate more changes in the direction, strength, content, and persistence of student behavior than will teachers who ignore this kind of communication.

(f) Children from the lower class, or handicapped children, or delinquent children, will have different concepts of adequacy and perceived-selves than children from contrasting categories and thus learning will be different for them.

(g) Children who see the acquisition of some skill as increasing congruency between the concept of adequacy and the perceived-self will, more rapidly than others, learn that skill on teaching machines.

(h) Negative methods by teachers (punishment, sarcasm, threats, etc.) will have less influence on concepts of adequacy than supportive methods.

(i) Academic-type objectives (like going to college) will not be part of becoming for some students even though the instrumental learning involved might be.

Summary

In the preceding discussion a self-concept theory of learning has been developed. The theory proceeds from three assumptions: (1) man is purposeful, (2) man is becoming, and (3) man is aware-ing. All behavior is seen as a process of expressing the self. An individual develops a concept of adequacy and a concept of perceived-self. To some extent these overlap, but in many ways they are discrepant. These discrepancies must be harmonized by the individual's striving to become adequate. This motivation takes the form of action as the individual defines what action is appropriate to his definition of adequacy. Consequences of these actions are evaluated and lead to modifications of behavior. This leads to the definition of learning as being the changing of relations between a self and its perceived-world as the self is expressed in striving to become adequate. A distinction is made between intrinsic and instrumental learning, one concerned with the self concepts and the direction of action and the other concerned with skills and the content of action. Implications of this theory are presented in a series of hypotheses about actual learning situations.

References

1. Allport, G. W. *Becoming.* New Haven: Yale University Press, 1955.

2. Bills, R. E., Vance, E. L., and Nelson, O. S. An index of adjustment and values. *J. Consult. Psychol.,* 1951, **15**, 257–261.
3. Cantril, H. *The Why of Men's Experience.* Macmillan, 1950.
4. Estes, W. K. Learning. In C. W. Harris (Ed.), *Encyclopedia of educational research* (3rd ed.). New York: Macmillan, 1960.
5. Fromm, E. *The Sane Society.* New York: Rinehart, 1955.
6. Goldstein, K. *The Organism.* New York: Amer. Book, 1939.
7. Hebb, D. O. The American revolution. *Amer. Psychologist,* 1960, **15**, 735–745.
8. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.,* 1931, **38**, 487–506.
9. Maslow, A. H. *Motivation and Personality.* New York: Harper, 1954.
10. Mowrer, O. H. *Learning Theory and Behavior.* New York: Wiley, 1960.
11. Olds, J. *The Growth and Structure of Motives.* Glencoe, Ill.: Free Press, 1956.
12. Orzeck, A. Z., McGuire, C., and Longenecker, E. D. Multiple self concepts as affected by mood states. *Amer. J. Psychiatry,* 1958, **115**, 349–353.
13. Rogers, C. R. *Client-Centered Therapy.* Boston: Houghton-Mifflin, 1951.
14. Rogers, C. R., and Dymond, Rosalind F. *Psychotherapy and Personality Change.* Chicago: Univ. of Chicago Press, 1954.
15. Sears, R. R. Social behavior and personality development. In Talcott Parsons & Edward A. Shils (Ed.), *Toward a General Theory of Action.* Cambridge, Mass.: Harvard Univer. Press, 1951.
16. White, R. W. Motivation reconsidered: the concept of competence. *Psychol. Rev.,* 1959, **66**, 297–333.
17. Wylie, Ruth. *The Self Concept.* Lincoln: Univer. of Nebraska Press, 1961.
18. Gilbran, K. *The Prophet* (pocket ed.). New York: Knopf, 1960.

4.3 *The Act of Discovery*

JEROME S. BRUNER

Reprinted from the *Harvard Educational Review,* 1961, **31,** 21–22, with permission of the author. Copyright © 1961 by the President and Fellows of Harvard College. Jerome S. Bruner's early research interests were in experimental and personality psychology, but in the last two decades he has become an international figure in the field of cognitive functioning and development.

In this article, Bruner argues in favor of a heuristic approach to learning and problem solving and makes the point that "material that is organized in terms of a person's own interests and cognitive structures is material that has the best chance of being accessible in memory." Although his terminology is different from that employed by Beatty and Clark in the preceding article, his viewpoint is quite compatible with a number of the principles put forth by them.

Maimonides, in his *Guide for the Perplexed,*[1] speaks of four forms of perfection that men might seek. The first and lowest form is perfection in the acquisition of worldly goods. The great philosopher dismisses such perfection on the ground that the possessions one acquires bear no meaningful relation to the possessor: "A great king may one morning find that there is no difference between him and the lowest person." A second perfection is of the body, its conformation and skills. Its failing is that it does not reflect on what is uniquely human about man: "he could [in any case] not be as strong as a mule." Moral perfection is the third, "the highest degree of excellency in man's character." Of this perfection Maimonides says: "Imagine a person being alone, and having no connection whatever with any other person; all his good moral principles are at rest, they are not required and give man no perfection whatever. These principles are only necessary and useful when man comes in contact with others." "The fourth kind of perfection is the true perfection of man; the possession of the highest intellectual faculties. . . ." In justification of his assertion, this extraordinary Spanish-Judaic philosopher urges: "Examine the first three kinds of perfection; you will find that if you possess them, they are not your property, but the property of others. . . . But the last kind of perfection is exclusively yours; no one else owns any part of it."

It is a conjecture much like that of Maimonides that leads me to examine the act of discovery in man's intellectual life. For if man's intellectual excellence is the most his own among his perfections, it is also the case that the most uniquely personal of all that he knows is that which he has discovered for himself. What difference does it make, then, that we encourage discovery in the learning of the young? Does it, as Maimonides would say, create a special and unique relation between knowledge possessed and the possessor? And what may such a unique relation do for a man—or for a child, if you will, for our concern is with the education of the young?

The immediate occasion for my concern with discovery—and I do not restrict discovery to the act of finding out something that before was unknown to mankind, but rather include all forms of obtaining knowledge for oneself by the use of one's own mind—the immediate occasion is the work of the various new curriculum projects that have grown up in America during the last six or seven years. For whether one speaks to mathematicians or physicists or historians, one encounters repeatedly an expression of faith in the powerful effects that come from permitting the student to put things together for himself, to be his own discoverer.

First, let it be clear what the act of discovery entails. It is rarely, on the frontier of knowledge or elsewhere, that new facts are "discovered" in the sense of being encountered as Newton suggested in the form of islands of truth in an uncharted sea of ignorance. Or if they appear to be discovered in this way, it is almost always thanks to some happy hypotheses about where to navigate. Discovery, like surprise,

favors the well prepared mind. In playing bridge, one is surprised by a hand with no honors in it at all and also by hands that are all in one suit. Yet all hands in bridge are equiprobable: one must know to be surprised. So too in discovery. The history of science is studded with examples of men "finding out" something and not knowing it. I shall operate on the assumption that discovery, whether by a schoolboy going it on his own or by a scientist cultivating the growing edge of his field, is in its essence a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights. It may well be that an additional fact or shred of evidence makes this larger transformation of evidence possible. But it is often not even dependent on new information.

It goes without saying that, left to himself, the child will go about discovering things for himself within limits. It also goes without saying that there are certain forms of child rearing, certain home atmospheres that lead some children to be their own discoverers more than other children. These are both topics of great interest, but I shall not be discussing them. Rather, I should like to confine myself to the consideration of discovery and "finding-out-for-oneself" within an educational setting—specifically the school. Our aim as teachers is to give our student as firm a grasp of a subject as we can, and to make him as autonomous and self-propelled a thinker as we can—one who will go along on his own after formal schooling has ended. I shall return in the end to the question of the kind of classroom and the style of teaching that encourages an attitude of wanting to discover. For purposes of orienting the discussion, however, I would like to make an overly simplified distinction between teaching that takes place in the *expository mode* and teaching that utilizes the *hypothetical mode.* In the former, the decisions concerning the mode and pace and style of exposition are principally determined by the teacher as expositor; the student is the listener. If I can put the matter in terms of structural linguistics, the speaker has a quite different set of decisions to make than the listener: the former has a wide choice of alternatives for structuring, he is anticipating paragraph content while the listener is still intent on the words, he is manipulating the content of the material by various transformations, while the listener is quite unaware of these internal manipulations. In the hypothetical mode, the teacher and the student are in a more cooperative position with respect to what in linguistics would be called "speaker's decisions." The student is not a bench-bound listener, but is taking a part in the formulation and at times may play the principal role in it. He will be aware of alternatives and may even have an "as if" attitude toward these and, as he receives information he may evaluate it as it comes. One cannot describe the process in either mode with great precision as to detail, but I think the foregoing may serve to illustrate what is meant.

Consider now what benefit might be derived from the experience of learning

through discoveries that one makes for oneself. I should like to discuss these under four headings: (1) The increase in intellectual potency, (2) the shift from extrinsic to intrinsic rewards, (3) learning the heuristics of discovering, and (4) the aid to memory processing.

1. Intellectual Potency. If you will permit me, I would like to consider the difference between subjects in a highly constrained psychological experiment involving a two-choice apparatus. In order to win chips, they must depress a key either on the right or the left side of the machine. A pattern of payoff is designed such that, say, they will be paid off on the right side 70 per cent of the time, on the left 30 per cent, although this detail is not important. What is important is that the payoff sequence is arranged at random, and there is no pattern. I should like to contrast the behavior of subjects who think that there *is* some pattern to be found in the sequence—who think that regularities are discoverable—in contrast to subjects who think that things are happening quite by *chance.* The former group adopts what is called an "event-matching" strategy in which the number of responses given to each side is roughly equal to the proportion of times it pays off: in the present case R70:L30. The group that believes there is no pattern very soon reverts to a much more primittive strategy wherein *all* responses are allocated to the side that has the greater payoff. A little arithmetic will show you that the lazy all-and-none strategy pays off more if indeed the environment is random: namely, they win seventy percent of the time. The event-matching subjects win about 70% on the 70% payoff side (or 49% of the time there) and 30% of the time on the side that pays off 30% of the time (another 9% for a total take-home wage of 58% in return for their labors of decision). But the world is not always or not even frequently random, and if one analyzes carefully what the event-matchers are doing, it turns out that they are trying out hypotheses one after the other, all of them containing a term such that they distribute bets on the two sides with a frequency to match the actual occurrence of events. If it should turn out that there is a pattern to be discovered, their payoff would become 100%. The other group would go on at the middling rate of 70%.

What has this to do with the subject at hand? For the person to search out and find regularities and relationships in his environment, he must be armed with an expectancy that there will be something to find and, once aroused by expectancy, he must devise ways of searching and finding. One of the chief enemies of such expectancy is the assumption that there is nothing one can find in the environment by way of regularity or relationship. In the experiment just cited, subjects often fall into a habitual attitude that there is either nothing to be found or that they can find a pattern by looking. There is an important sequel in behavior to the two attitudes, and to this I should like to turn now.

We have been conducting a series of experimental studies on a group of some seventy school children over the last four years. The studies have led

us to distinguish an interesting dimension of cognitive activity that can be described as ranging from *episodic empiricism* at one end to *cumulative constructionism* at the other. The two attitudes in the choice experiments just cited are illustrative of the extremes of the dimension. I might mention some other illustrations. One of the experiments employs the game of Twenty Questions. A child—in this case he is between 10 and 12—is told that a car has gone off the road and hit a tree. He is to ask questions that can be answered by "yes" or "no" to discover the cause of the accident. After completing the problem, the same task is given him again, though he is told that the accident had a different cause this time. In all, the procedure is repeated four times. Children enjoy playing the game. They also differ quite markedly in the approach or strategy they bring to the task. There are various elements in the strategies employed. In the first place, one may distinguish clearly between two types of questions asked: the one is designed for locating constraints in the problem, constraints that will eventually give shape to an hypothesis; the other is the hypothesis as question. It is the difference between, "Was there anything wrong with the driver?" and "Was the driver rushing to the doctor's office for an appointment and the car got out of control?" There are children who precede hypotheses with efforts to locate constraint and there are those who, to use our local slang, are "pot-shotters," who string out hypotheses non-cumulatively one after the other. A second element of strategy is its connectivity of information gathering: the extent to which questions asked utilize or ignore or violate information previously obtained. The questions asked by children tend to be organized in cycles, each cycle of questions usually being given over to the pursuit of some particular notion. Both within cycles and between cycles one can discern a marked difference on the connectivity of the child's performance. Needless to say, children who employ constraint location as a technique preliminary to the formulation of hypotheses tend to be far more connected in their harvesting of information. Persistence is another feature of strategy, a characteristic compounded of what appear to be two components: a sheer doggedness component, and a persistence that stems from the sequential organization that a child brings to the task. Doggedness is probably just animal spirits or the need for achievement—what has come to be called *n-ach.* Organized persistence is a maneuver for protecting our fragile cognitive apparatus from overload. The child who has flooded himself with disorganized information from unconnected hypotheses will become discouraged and confused sooner than the child who has shown a certain cunning in his strategy of getting information—a cunning whose principal component is the recognition that the value of information is not simply in getting it but in being able to carry it. The persistence of the organized child stems from his knowledge of how to organize questions in cycles, how to summarize things to himself, and the like.

Episodic empiricism is illustrated by

information gathering that is unbound by prior constraints, that lacks connectivity, and that is deficient in organizational persistence. The opposite extreme is illustrated by an approach that is characterized by constraint sensitivity, by connective maneuvers, and by organized persistence. Brute persistence seems to be one of those gifts from the gods that make people more exaggeratedly what they are*

Before returning to the issue of discovery and its role in the development of thinking, let me say a word more about the ways in which information may get transformed when the problem solver has actively processed it. There is first of all a pragmatic question: what does it take to get information processed into a form best designed to fit some future use? Take an experiment by Zajonc[2] as a case in point. He gives groups of subjects information of a controlled kind, some groups being told that their task is to transmit the information to others, others that it is merely to be kept in mind. In general, he finds more differentiation and organization of the information received with the intention of being transmitted than there is for information received passively. An active set leads to a transformation related to a task to be performed. The risk, to be sure, is in possible overspecialization of information processing that may lead to such a high degree of specific organization that information is lost for general use.

I would urge now in the spirit of an hypothesis that emphasis upon discovery in learning has precisely the effect upon the learner of leading him to be a constructionist, to organize what he is encountering in a manner not only designed to discover regularity and relatedness, but also to avoid the kind of information drift that fails to keep account of the uses to which information might have to be put. It is, if you will, a necessary condition for learning the variety of techniques of problem solving, of transforming information for better use, indeed for learning how to go about the very task of learning. Practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving. So goes the hypothesis. It is still in need of testing. But it is an hypothesis of such important human implications that we cannot afford not to test it—and testing will have to be in the schools.

2. Intrinsic and Extrinsic Motives. Much of the problem in leading a child to effective cognitive activity is to free him from the immediate control of environmental rewards and punishments. That is to say, learning that starts in response to the rewards of parental or teacher approval or the avoidance of failure can too readily develop a pattern in which the child is seeking cues as to how to conform to what is expected of him. We know from studies of children who tend to be early over-achievers in school that

*I should also remark in passing that the two extremes also characterize concept attainment strategies as reported in *A Study of Thinking* by J. S. Bruner *et al.* (New York: John Wiley, 1956). Successive scanning illustrates well what is meant here by episodic empiricism; conservative focussing is an example of cumulative constructionism.

they are likely to be seekers after the "right way to do it" and that their capacity for transforming their learning into viable thought structures tends to be lower than children merely achieving at levels predicted by intelligence tests. Our tests on such children show them to be lower in analytic ability than those who are not conspicuous in over-achievement.[3] As we shall see later, they develop rote abilities and depend upon being able to "give back" what is expected rather than to make it into something that relates to the rest of their cognitive life. As Maimonides would say, their learning is not their own.

The hypothesis that I would propose here is that to the degree that one is able to approach learning as a task of discovering something rather than "learning about" it, to that degree will there be a tendency for the child to carry out his learning activities with the autonomy of self-reward or, more properly by reward that is discovery itself.

To those of you familiar with the battles of the last half-century in the field of motivation, the above hypothesis will be recognized as controversial. For the classic view of motivation in learning has been, until very recently, couched in terms of a theory of drives and reinforcement: that learning occurred by virtue of the fact that a response produced by a stimulus was followed by the reduction in a primary drive state. The doctrine is greatly extended by the idea of secondary reinforcement: any state associated even remotely with the reduction of a primary drive could also have the effect of producing learning. There has recently appeared a most searching and important criticism of this position, written by Professor Robert White,[4] reviewing the evidence of recently published animal studies, of work in the field of psychoanalysis, and of research on the development of cognitive processes in children. Professor White comes to the conclusion, quite rightly I think, that the drive-reduction model of learning runs counter to too many important phenomena of learning and development to be either regarded as general in its applicability or even correct in its general approach. Let me summarize some of his principal conclusions and explore their applicability to the hypothesis stated above.

I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence. According to Webster, competence means fitness or ability, and the suggested synonyms include capability, capacity, efficiency, proficiency, and skill. It is therefore a suitable word to describe such things as grasping and exploring, crawling and walking, attention and perception, language and thinking, manipulating and changing the surroundings, all of which promote an effective—a competent—interaction with the environment. It is true of course, that maturation plays a part in all these developments, but this part is heavily overshadowed by learning in all the more complex accomplishments like speech or skilled manipulation. I shall argue that it is necessary to make competence a mo-

tivational concept; there is *competence motivation* as well as competence in its more familiar sense of achieved capacity. The behavior that leads to the building up of effective grasping, handling, and letting go of objects, to take one example, is not random behavior that is produced by an overflow of energy. It is directed, selective, and persistent, and it continues not because it serves primary drives, which indeed it cannot serve until it is almost perfected, but because it satisfies an intrinsic need to deal with the environment.[5]

I am suggesting that there are forms of activity that serve to enlist and develop the competence motive, that serve to make it the driving force behind behavior. I should like to add to White's general premise that the *exercise* of competence motives has the effect of strengthening the degree to which they gain control over behavior and thereby reduce the effects of extrinsic rewards or drive gratification.

The brilliant Russian psychologist Vigotsky[6] characterizes the growth of thought processes as starting with a dialogue of speech and gesture between child and parent; autonomous thinking begins at the stage when the child is first able to internalize these conversations and "run them off" himself. This is a typical sequence in the development of competence. So too in instruction. The narrative of teaching is of the order of the conversation. The next move in the development of competence is the internalization of the narrative and its "rules of generation" so that the child is now capable of running off the narrative on his own. The hypothetical mode in teaching by encouraging the child to participate in "speaker's decisions" speeds this process along. Once internalization has occurred, the child is in a vastly improved position from several obvious points of view—notably that he is able to go beyond the information he has been given to generate additional ideas that can either be checked immediately from experience or can, at least, be used as a basis for formulating reasonable hypotheses. But over and beyond that, the child is now in a position to experience success and failure not as reward and punishment, but as information. For when the task is his own rather than a matter of matching environmental demands, he becomes his own paymaster in a certain measure. Seeking to gain control over his environment, he can now treat success as indicating that he is on the right track, failure as indicating he is on the wrong one.

In the end, this development has the effect of freeing learning from immediate stimulus control. When learning in the short run leads only to pellets of this or that rather than to mastery in the long run, then behavior can be readily "shaped" by extrinsic rewards. When behavior becomes more long-range and competence-oriented, it comes under the control of more complex cognitive structures, plans and the like, and operates more from the inside out. It is interesting that even Pavlov, whose early account of the learning process was based entirely on a notion of stimulus control of behavior through the conditioning mechanism in which, through contiguity a new conditioned

stimulus was substituted for an old unconditioned stimulus by the mechanism of stimulus substitution, that even Pavlov recognized his account as insufficient to deal with higher forms of learning. To supplement the account, he introduced the idea of the "second signalling system," with central importance placed on symbolic systems such as language in mediating and giving shape to mental life. Or as Luria[7] has put it, "the first signal system [is] concerned with directly perceived stimuli, the second with systems of verbal elaboration." Luria, commenting on the importance of the transition from first to second signal system, says: "It would be mistaken to suppose that verbal intercourse with adults merely changes the contents of the child's conscious activity without changing its form. . . . The word has a basic function not only because it indicates a corresponding object in the external world, but also because it abstracts, isolates the necessary signal, generalizes perceived signals and relates them to certain categories; it is this systematization of direct experience that makes the role of the word in the formation of mental processes so exceptionally important."[8,9]

It is interesting that the final rejection of the universality of the doctrine of reinforcement in direct conditioning came from some of Pavlov's own students. Ivanov-Smolensky[10] and Krasnogorsky[11] published papers showing the manner in which symbolized linguistic messages could take over the place of the unconditioned stimulus and of the unconditioned response (gratification of hunger) in children. In all instances, they speak of these as *replacements* of lower, first-system mental or neural processes by higher order or second-system controls. A strange irony, then, that Russian psychology that gave us the notion of the conditioned response and the assumption that higher order activities are built up out of colligations or structurings of such primitive units, rejected this notion while much of American learning psychology has stayed until quite recently within the early Pavlovian fold (see, for example, a recent article by Spence[12] in the *Harvard Educational Review* or Skinner's treatment of language[13] and the attacks that have been made upon it by linguists such as Chomsky[14] who have become concerned with the relation of language and cognitive activity). What is the more interesting is that Russian pedagogical theory has become deeply influenced by this new trend and is now placing much stress upon the importance of building up a more active symbolical approach to problem solving among children.

To sum up the matter of the control of learning, then, I am proposing that the degree to which competence or mastery motives come to control behavior, to that degree the role of reinforcement or "extrinsic pleasure" wanes in shaping behavior. The child comes to manipulate his environment more actively and achieves his gratification from coping with problems. Symbolic modes of representing and transforming the environment arise and the importance of stimulus-response-

reward sequences declines. To use the metaphor that David Riesman developed in a quite different context, mental life moves from a state of outer-directedness in which the fortuity of stimuli and reinforcement are crucial to a state of inner-directedness in which the growth and maintenance of mastery become central and dominant.

3. Learning the Heuristics of Discovery. Lincoln Steffens,[15] reflecting in his *Autobiography* on his undergraduate education at Berkeley, comments that his schooling was overly specialized in learning about the known and that too little attention was given to the task of finding out about what was not known. But how does one train a student in the techniques of discovery? Again I would like to offer some hypotheses. There are many ways of coming to the arts of inquiry. One of them is by careful study of its formalization in logic, statistics, mathematics, and the like. If a person is going to pursue inquiry as a way of life, particularly in the sciences, certainly such study is essential. Yet, whoever has taught kindergarten and the early primary grades or has had graduate students working with him on their theses —I choose the two extremes for they are both periods of intense inquiry— knows that an understanding of the formal aspect of inquiry is not sufficient. There appear to be, rather, a series of activities and attitudes, some directly related to a particular subject and some of them fairly generalized, that go with inquiry and research. These have to do with the *process* of trying to find out something and while they provide no guarantee that the *product* will be any *great* discovery, their absence is likely to lead to awkwardness or aridity or confusion. How difficult it is to describe these matters —the heuristics of inquiry. There is one set of attitudes or ways of doing that has to do with sensing the relevance of variables—how to avoid getting stuck with edge effects and getting instead to the big sources of variance. Partly this gift comes from intuitive familiarity with a range of phenomena, sheer "knowing the stuff." But it also comes out of a sense of what things among an ensemble of things "smell right" in the sense of being of the right order of magnitude or scope or severity.

The English philosopher Weldon describes problem solving in an interesting and picturesque way. He distinguishes between difficulties, puzzles, and problems. We solve a problem or make a discovery when we impose a puzzle form on to a difficulty that converts it into a problem that can be solved in such a way that it gets us where we want to be. That is to say, we recast the difficulty into a form that we know how to work with, then work it. Much of what we speak of as discovery consists of knowing how to impose what kind of form on various kinds of difficulties. A small part but a crucial part of discovery of the highest order is to invent and develop models or "puzzle forms" that can be imposed on difficulties with good effect. It is in this area that the truly powerful mind shines. But it is interesting to what

degree perfectly ordinary people can, given the benefit of instruction, construct quite interesting and what, a century ago, would have been considered greatly original models.

Now to the hypothesis. It is my hunch that it is only through the exercise of problem solving and the effort of discovery that one learns the working heuristic of discovery, and the more one has practice, the more likely is one to generalize what one has learned into a style of problem solving or inquiry that serves for any kind of task one may encounter—or almost any kind of task. I think the matter is self-evident, but what is unclear is what kinds of training and teaching produce the best effects. How do we teach a child to, say, cut his losses but at the same time be persistent in trying out an idea; to risk forming an early hunch without at the same time formulating one *so* early and with so little evidence as to be stuck with it waiting for appropriate evidence to materialize; to pose good testable guesses that are neither too brittle nor too sinuously incorrigible; etc., etc. Practice in inquiry, in trying to figure out things for oneself is indeed what is needed, but in what form? Of only one thing I am convinced, I have never seen anybody improve in the art and technique of inquiry by any means other than engaging in inquiry.

4. Conservation of Memory. I should like to take what some psychologists might consider a rather drastic view of the memory process. It is a view that in large measure derives from the work of my colleague, Professor George Miller.[16] Its first premise is that the principal problem of human memory is not storage, but retrieval. In spite of the biological unlikeliness of it, we seem to be able to store a huge quantity of information—perhaps not a full tape recording, though at times it seems we even do that, but a great sufficiency of impressions. We may infer this from the fact that recognition (i.e., recall with the aid of maximum prompts) is so extraordinarily good in human beings—particularly in comparison with spontaneous recall where, so to speak, we must get out stored information without external aids or prompts. The key to retrieval is organization or, in even simpler terms, knowing where to find information and how to get there.

Let me illustrate the point with a simple experiment. We present pairs of words to twelve-year-old children. One group is simply told to remember the pairs, that they will be asked to repeat them later. Another is told to remember them by producing a word or idea that will tie the pair together in a way that will make sense to them. A third group is given the mediators used by the second group when presented with the pairs to aid them in tying the pairs into working units. The word pairs include such juxtapositions as "chair-forest," "sidewalk-square," and the like. One can distinguish three styles of mediators and children can be scaled in terms of their relative preference for each: *generic mediation* in which a pair is tied together by a superordinate idea: "chair and forest are both made of wood"; *thematic*

mediation in which the two terms are imbedded in a theme or little story: "the lost child sat on a chair in the middle of the forest"; and *part-whole mediation* where "chairs are made from trees in the forest" is typical. Now, the chief result, as you would all predict, is that children who provide their own mediators do best—indeed, one time through a set of thirty pairs, they recover up to 95% of the second words when presented with the first ones of the pairs, whereas the uninstructed children reach a maximum of less than 50% recovered. Interestingly enough, children do best in recovering materials tied together by the form of mediator they most often use.

One can cite a myriad of findings to indicate that any organization of information that reduces the aggregate complexity of material by imbedding it into a cognitive structure a person has constructed will make that material more accessible for retrieval. In short, we may say that the process of memory, looked at from the retrieval side, is also a process of problem solving: how can material be "placed" in memory so that it can be got on demand?

We can take as a point of departure the example of the children who developed their own technique for relating the members of each word pair. You will recall that they did better than the children who were given by exposition the mediators they had developed. Let me suggest that in general, material that is organized in terms of a person's own interests and cognitive structures is material that has the best chance of being accessible in memory. That is to say, it is more likely to be placed along routes that are connected to one's own ways of intellectual travel.

In sum, the very attitudes and activities that characterize "figuring out" or "discovering" things for oneself also seem to have the effect of making material more readily accessible in memory.

References

1. Maimonides, *Guide for the Perplexed* (New York: Dover Publications, 1956).
2. R. B. Zajonc (Personal communication, 1957).
3. J. S. Bruner and A. J. Caron, "Cognition, Anxiety, and Achievement in the Pre-adolescent," *Journal of Educational Psychology* (in press).
4. R. W. White, "Motivation Reconsidered: The Concept of Competence," *Psychological Review,* LXVI (1959), 297–333.
5. *Ibid.,* pp. 317–18.
6. L. S. Vigotsky, *Thinking and Speech* (Moscow, 1934).
7. A. L. Luria, "The Directive Function of Speech in Development and Dissolution," *Word,* XV (1959), 341–464.
8. *Ibid.,* p. 12.
9. For an elaboration of the view expressed by Luria, the reader is referred to the forthcoming translation of L. S. Vigotsky's 1934 book being published by John Wiley and Sons and the Technology Press.
10. A. G. Ivanov-Smolensky, "Concerning the Study of the Joint Activity of the First and Second Signal Systems," *Journal of Higher Nervous Activity,* I (1951), 1.
11. N. D. Krasnogorsky, *Studies of Higher Nervous Activity in Animals and in Man,* Vol. I (Moscow, 1954).
12. K. W. Spence, "The Relation of Learning Theory to the Technique of Education,"

Harvard Educational Review, XXIX (1959), 84–95.

13. B. F. Skinner, *Verbal Behavior* (New York: Appleton-Century-Crofts, 1957).
14. N. Chomsky, *Syntactic Structure* (The Hague, The Netherlands: Mouton & Co., 1957).
15. L. Steffens, *Autobiography of Lincoln* Steffens (New York: Harcourt, Brace, 1931).
16. G. A. Miller, "The Magical Number Seven, Plus or Minus Two," *Psychological Review,* LXIII (1956), 81–97.

4.4 *The Motivating Effect of Learning by Directed Discovery*

BERT Y. KERSH

Reprinted from the *Journal of Educational Psychology,* 1962, **53,** 65–71, with permission of the author and the American Psychological Association, Inc. Bert Y. Kersh is the dean of the faculty at the Oregon College of Education in Monmouth.

Kersh's research study continues the theme of learning through discovery that was discussed by Bruner in the preceding article. Kersh found that directed discovery produced more favorable motivation on the part of students, but did not necessarily lead to superior results. It evidently works best when it is tried over a period of time and not in a brief, one-trial experience like that used in this experiment. Later research (not reported in this article, of course) suggests that teacher-directed instruction tends to produce better short-term gains, but that discovery approaches give better retention weeks and months after the learning experience was completed.

Advocates of the process of learning by directed discovery claim a number of advantages, most of which are included in a recent article by Bruner (1961). He has suggested that learning by discovery benefits the learner in four ways: it (*a*) increases the learner's ability to learn related material, (*b*) fosters an interest in the activity itself rather than in the rewards which may follow from the learning, (*c*) develops ability to approach problems in a way that will more likely lead to a solution, and (*d*) tends to make the material that is learned easier to retrieve or reconstruct.

Research evidence does not entirely support Bruner's arguments. One of the more recent reviewers, Ausubel (1961, p. 47), concludes "that most of the reasonably well-controlled studies report negative findings." However, as is true in other areas of research, the evidence is somewhat equivocal, partly because it is difficult to equate studies in terms of the amount and kind of direction that is provided. The experimental subjects rarely if ever are required to learn completely without help, and the kind of help provided commonly differs. Consequently, there are studies which appear to be somewhat contradictory, such as Craig's (1956), in which the "directed" group learned and retained significantly more principles than the "independent" group, and Kittell's (1957), in which the group which received an intermediate amount of guidance was superior in learning, retention, and transfer to groups receiving either more or less direction. It has been suggested that the "intermediate" amount of guidance provided by Kittell may have actually exceeded the amount Craig provided to his directed group (Ausubel, 1961, p. 52).

One of the few studies that forced learners to discover almost entirely without help provides data in support of the discovery process (Kersh, 1958). The contrasting directed treatment groups were superior in learning rate and immediate recall, but the "no help" group was superior in terms of retention and transfer after a period of approximately one month following the learning period. No evidence was produced to indicate that the no help group understood the rules better. Instead, an explanation was offered in terms of practice. On the basis of a subjective analysis of the subjects' comments written on the retests and reported to the experimenter, it was concluded that the learners were motivated to continue the learning process or to continue practicing the task after the formal learning period.

The present experiment was designed to provide formal data concerning the motivating power in question.

Hypothesis

Each subject had the task of learning the following two rules of addition:

1. Odd Numbers rule. The sum of any series of consecutive odd numbers beginning with 1 is equal to the square of the number of figures in the series. (For example, 1, 3, 5, 7, is such a series; there are four numbers, so 4×4 is 16, the sum.)
2. Constant Difference rule. The sum of any series of numbers in which the

difference between the numbers is constant is equal to one-half the product of the number of figures and the sum of the first and last numbers. (For example, 2, 3, 4, 5, is such a series; 2 and 5 are 7; therefore; there are four figures, so 4 × 7 is 28; half of 28 is 14 which is the sum.)

The rules can be learned by simple memorization of the task procedure as above. Further, the learner can become cognizant of certain relations which these rules bear to geometrical and arithmetical concepts, in which case it is assumed that his learning will be more meaningful. The definition of meaning, as well as the geometrical and arithmetical relationships referred to are identified in a previous publication (Kersh, 1958). In the hypothesis statement below, the term "relationships" refers specifically to those in the reference cited above and generally to comparable relationships in related tasks.

As will be explained below, the experimental treatments in the present study differed primarily with respect to the extent of the external direction provided the subjects in learning the relationships referred to above. The present experiment was designed to test the following hypothesis.

To the extent that the external direction provided to the learner is lessened during his attempts to discover the relationships which are considered essential to the understanding of a cognitive task: (*a*) the learner will tend to use the learned material more frequently after the learning period (i.e., to extend the practice period voluntarily) and, as a result, (*b*) he will remember it longer and transfer his learning more effectively.

It should be noted that the hypothesis is written in two parts and that the second is dependent upon the first.

Procedure

A total of 90 high school geometry students was utilized, having been selected from a larger group on the basis of a pretest covering the arithmetical and geometrical concepts and procedures that were considered essential prerequisites to the tasks used in the experiment. The entire sample was then taught the two rules of addition given above by being simply told the rules and given practice in their application. They were taught by a programed booklet procedure to the same criterion, six successive applications of each of the two rules. Thereafter, the subjects were divided at random into three main groups of 30 each, and each group was treated differently.

One group, called the Directed Learning group, was taught the rules and their explanation entirely by a programed learning technique. Each subject learned from a booklet in which the learning process was broken down into smaller steps, and answers to questions or solutions to problems were revealed to the subject whether he responded correctly or not.

A second group, called the Guided Discovery group, was required to discover the explanation with guidance from the experimenter. The subjects in the Guided Discovery group were taught tutorially using a form of Socrat-

ic questioning which required each subject to perform specific algebraic manipulations and to make inferences without help. The guidance was a practical expedient, since it was necessary to control between groups the quality and quantity of the relationships used in explaining the rules.

The final group was called, appropriately, the Rote Learning group since the explanation for the rules was omitted. This treatment was incorporated in the research design primarily as the control for "meaningful" learning.

After the learning period of the experiment, a test of recall and transfer was given to subgroups of each treatment group after 3 days, 2 weeks, and 6 weeks. For this purpose each of the three main groups was divided into three subgroups of 10 each.

The test consisted of two problems and a short questionnaire. The problems were given first with instructions to show all work including scratchwork. The two test problems were as follows:

1. John's employer agrees to pay him $1.00 for his first day of work and increase his pay by $2.00 each day. How much will he receive for the first month's work if he works all 30 days?
2. A man is left a sum of money by an eccentric relative. The will states that he will receive $10.00 the first month and that each successive monthly payment will be increased by $5.00 (i.e., he will receive $10.00 the first month, $15.00 the second month, $20.00 the third month, etc.). His monthly payment at the end of four years is $245.00. What is the total amount he has been paid by that time?

The questionnaire asked the subject to state each rule, using examples if necessary, and to report whether or not he made use of the rules after the formal learning period.

Results

The number of subjects in each group who used the appropriate rule in an acceptable way on the test was employed as the index of transfer. Acceptable use of a rule for the first test problem meant the use of either rule to obtain the solution; for the second test problem, only the Constant Difference rule was acceptable. Computational accuracy was not required.

The number of subjects in each group who wrote an acceptable statement of each rule was used as a measure of pure retention. To be acceptable, each subject's statement had to be complete and accurate, but not necessarily in the same words as the original statements. Errors in spelling or grammar were overlooked.

Table 1 presents the number who used and stated the rules in the acceptable way on the test problems. A total of 90 subjects served as the basis for the data in Table 1, 10 subjects per cell.

In the statistical analysis, use was made of a chi-square technique devised by Li (1957, pp. 416–420). The data included under each of the columns of Table 1 were envisioned as a 2×9 contingency table, with 8 *df.*

TABLE 1. Number of Subjects (of 10 in Each Cell) Who Used and Stated the Rules Correctly on the Retest

Treatment Groups	Used Rules		Stated Rules	
	Odd Numbers [a] (1)	Constant Difference (2)	Odd Numbers [a] (3)	Constant Difference [a] (4)
Rote learning:				
3 days	7	7	7	9
2 weeks	7	6	2	6
6 weeks	4	4	0	3
Guided discovery:				
3 days	6	6	8	9
2 weeks	3	5	3	4
6 weeks	2	3	3	3
Directed learning:				
3 days	4	3	3	4
2 weeks	4	3	1	3
6 weeks	0	3	1	1

[a] Differences between treatment groups and between test periods significant by chi square at or beyond .05 level.

Four separate analyses were then conducted, each of which broke down the chi square into the following components: (*a*) differences between teaching treatments (2 *df*), (*b*) differences between test periods (2 *df*), and (*c*) differences attributed to interaction of treatments and time periods (4 *df*).

None of the interaction effects was significant, indicating that the rate of forgetting did not differ significantly across the teaching treatment groups. A trend analysis of the test data indicated also that the rate of forgetting was constant for all groups (Li, 1957, pp. 226–233).

Otherwise, as pointed out by the footnote references in Table 1, the differences between treatment groups and between test periods were found to be significant for all columns except that headed "Constant Difference 2," for which the observed differences were found not to be reliable.

Perhaps the most striking finding in the present study is that the Rote Learning group was found to be consistently superior in every respect to the other treatment groups. Although this completely unanticipated finding has no direct bearing on the hypothesis in question, it does nevertheless bear clearly upon the related question of "meaningful vs. mechanical" learning. This finding will be discussed in a subsequent section.

Strictly speaking, the hypothesis which the present experiment was designed to test involves only the Guided Discovery and Directed Learning treatments. To support the major hypothesis, the data should have shown that the subjects comprising the Guided Discovery group used the rules after

the learning period more frequently than the subjects in the Directed Learning group; and, if so, that the former remembered and transferred the rules more effectively than the latter.

With respect to the frequency of using the rules after the learning period, the results do support the hypothesis. Although the number of subjects in each group who reported that they did use the rules was very small, the difference between the frequency patterns of the two groups in question is statistically significant. Eleven subjects of the 30 in the Guided Discovery group reported that they had used the rules as compared with two subjects in the Directed Learning group. In the Rote Learning group, six subjects of 30 reported in the affirmative.

With respect to the relative permanence of the retention and increased transfer effects, the results also support the hypothesis. The Guided Discovery group is clearly superior to the Directed Learning group 3 days after the learning period, and since the rate of forgetting may be presumed to be approximately the same for each treatment group (see statistical analysis above), their initial superiority remains after 6 weeks.

Discussion

The data from this present experiment do not support the generalization that learning by a process which involves discovery is necessarily superior to learning by more highly directed processes. Indeed, these data suggest that under certain conditions of learning, highly formalized "lecture-drill" techniques, ordinarily considered sterile and meaningless, produce better results than techniques which attempt to develop "understanding."

One explanation for the present results is that they reflect a simple and well known phenomenon, retroactive inhibition. The experimental efforts to inject meaning into the rules amounted to following their initial rote learning with a closely related and complex learning task; thus the Rote Learning group may have surpassed other groups simply because retention among the latter was inhibited by the interpolated learning.

How may the present results be reconciled with those of the previous experiment by Kersh (1958), in which learning by discovery proved markedly superior? The preferred interpretation is that the findings of the two studies are actually complementary. Schematically, the treatments employed in the two experiments may be compared on a line representing the continuum of learning processes: at one extreme, learning without any external direction whatsoever (true self-discovery); at the other, learning by lecture-drill processes (rote learning), as follows:

1958 EXPERIMENT	No Help	Direct Reference		Rule Given
EXPERIMENT PRESENT		Guided Discovery	Directed Learning	Rote Learning

As is indicated above, the Direct Reference treatment in the 1958 experiment is comparable to the Guided Discovery treatment in the present one; similarly, the Rule Given and Rote Learning groups correspond. The present experiment has no counterpart to the No Help treatment of the previous study; and, in the previous one, the Directed Learning treatment was not represented.

When compared as above, the results of the two experiments are remarkably similar. The initial achievement of the comparable groups in both experiments was very high then dropped to where only about half of each group was able to recall and apply the rules after 4 to 6 weeks. In each experiment the difference in the performance of the Rote Learning and Directed Discovery groups was not notable; if anything, the Rote Learning groups tended to perform slightly better.

With respect to the motivating power of learning by discovery, in the 1958 experiment the superior performance of the No Help subjects on the retest together with their written comments and verbal reports to the experimenter strongly evidenced their increased interest. The present results leave no doubt that there is a tendency for interest to accrue as a result of learning by discovery.

The results of both experiments also are consistent in their failure to support the notion that attempts to provide added meaning will necessarily prolong memory for rules and procedures and will enhance their transfer. On the contrary, both experiments suggest that such attempts may well do more to interfere with learning than enhance it. This does not mean that rote learning is superior to learning with understanding. Rather it means that we need to know much more than we do about meaningful learning and how we come by it.

The relatively poor showing of the Directed Learning group in the present study is partially explained by the subjects' reported failure to practice the rules after the learning period to the extent that the subjects did in other groups. Why the Rote Learning treatment generated more interest than the treatment in question again may reflect nothing more than that the original learning was inhibited by the interpolated programed learning. The subjects' unfamiliarity with the instructional procedure may have contributed to their confusion.

Most certainly the data from the two experiments under discussion suggest that the frequently taught principles of learning that pertain to self-discovery and meaning (see introduction) should be restated or qualified. The following statements are offered for further study.

Learning by self-discovery. Learning by self-discovery is superior to learning with external direction only insofar as it increases student motivation to pursue the learning task. If sufficiently motivated, the student may then continue the learning process autonomously beyond the formal period of learning. As a result of his added experience, the learner may then raise his level of achievement, remember what he learned longer, and transfer it more effectively. The explanation for the elusive

drive generated by independent discovery is not evident, but several have been offered, including the Zeigarnik effect of superior memory for unfinished tasks and the Ovsiankina effect of resumption of incomplete tasks.[2] It also could be explained in terms of operant conditioning; specifically, as a kind of "searching behavior" reinforced by the experimenter's comments and by the subject's own successful progress toward a solution. Whatever the explanation, the motivating power evidently does not appear in strength unless the student is required to learn almost completely without help and expends intensive effort over a period of 15 minutes or more.

Meaningful learning. Aside from the advantage the student may come to have academically, he may not benefit from knowing the explanations for rules and procedures he learns, i.e., the pattern of relationships involved. That which is meaningful (understood) may or may not be retained longer and transferred more effectively than that which has been learned by rote. Moreover, superficial efforts to gain understanding after a rule or principle has been memorized may have an inhibitory effect when the student attempts to recall and transfer the original learning. If it is important only that the task be understood (as is most often the case, presumably), the essential relationships may be learned most economically when taught by another person or teaching program, not by process of self-discovery.

[2] The author is particularly indebted to Julius M. Sassenrath and the late Percival M. Symonds for their critical comments and suggestions.

Summary

High school students were taught 2 novel rules of addition by a programed booklet procedure. Subsequently, ⅓ of the 90 Ss were given individual guidance in discovering the explanation for the rules ("guided discovery"), ⅓ were taught the explanation by a programed booklet ("directed learning"), and the remaining ⅓ were given no further instruction ("rote learning"). A questionnaire and a test of recall and transfer given 3 days, 2 weeks, and 6 weeks later favored the Rote Learning and Guided Discovery groups. The questionnaire indicated that the Guided Discovery group practiced the rules during the time interval between the learning and test period more than Ss in other groups (chi square, $p = .05$). The data support the hypothesis that self-discovery motivates the S to practice more and thus to remember and transfer more than he might if taught directly.

References

Ausubel, D. P. Learning by discovery: Rationale and mystique. *Bull. Nat. Ass. Second. Sch. Principals,* 1961, **45**, 18–58.

Bruner, J. S. The act of discovery. *Harv. Educ. Rev.,* 1961, **31**, 21–32.

Craig, R. C. Directed versus independent discovery of established relations. *J. Educ. Psychol.,* 1956, **47**, 223–234.

Kersh, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. Educ. Psychol.,* 1958, **49**, 282–292.

Kittell, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *J. Educ., Psychol.,* 1957, **48**, 391–405.

Li, J. C. R. *Introduction to Statistical Inference,* Ann Arbor, Mich.: Edwards, 1957.

4.5 *A Comparative Study of Three Preschool Curricula*

DAVID P. WEIKART

Unpublished paper delivered before the biennial meeting of the Society for Research in Child Development, Santa Monica, California, March 1969, and included in this book of readings with the permission of the author. David P. Weikart directs the High Scope Educational Research Foundation and the research and development program of public schools of Ypsilanti, Michigan.

Weikart's paper describes the Ypsilanti Preschool Curriculum Research Project, a continuing experiment designed to compare three kinds of preschool experiences for socially disadvantaged children: a conventional, permissive nursery school; a cognitively oriented curriculum based on Piaget's theories; and a language training program based on concepts developed by Carl Bereiter at the University of Illinois. All three types of treatment produced somewhat similar cognitive gains, which in turn were superior to those attained by contrast groups. A subsequent follow-up (reported in the addendum) showed that the performance of the nursery-school group tended to lag behind that of the others. Weikart explains this finding not in terms of curricular differences but rather in terms of the fact that it is easier for teachers in cognitive and language programs to maintain performance levels.

This study is included here not only because it presents an interesting experiment, but also because the author describes the rationale for various types of experiments designed to help socially disadvantaged children. He also explores several "background factors" that significantly affect the kind of progress children make in school.

Introduction

Recent efforts through massive Federal and State programs to improve the intellectual and academic performance of disadvantaged children through compensatory education projects have met with only limited success. Both professionals and parents are gradually becoming aware of this fact, and the results are disquieting. Professionals are openly pessimistic and are often ready to reduce support for programs and limit their own involvement. Parents of disadvantaged children are increasingly aggressive in their demands for action. They are determined that someone assume the blame for the years of waste of human potential while ivory tower theorists wrestled with minutia and the education "establishment" promoted its own interest. And the worst is yet to come as the limitations of current compensatory education programs become more widely known.

Professional educators and psychologists are expressing in public observations and conclusions limited to private conversations several years ago. Jencks (1968) has pointed out in the New York Times Magazine that "Unfortunately, none of these (compensatory or remedial) programs has proved consistently successful over any significant period." This indictment has been verified by a group of researchers from the American Research Institute who reveiwed compensatory programs covering preschool through 12th grade for the period 1963 to 1968; Hawkridge, Chalupsky, and Roberts (1968) found only 21 compensatory educational programs which met a criterion of improved intellectual or academic functioning in a total sample of over 1000 such projects nominated for the study from throughout the country.

Parents in all minority groups are becoming increasingly outspoken in their criticism of the schools. Recently in Harlem a parent at a meeting of the advisory committee for a Follow Through program asked when the experimental project was going to teach her kindergarten youngster his ABC's. Strongly supported by the other parents with vigorous head nodding and such comments as "you tell it" and "that's right," the mother proceeded to read off the school in which her children had failed to learn and to place the blame on the teachers, the curriculum, and the administration. This parent had no concern for the theoretical basis of curriculum development; she wanted education for her child *now.* She was especially bitter about the implication, common in professional circles, that her children had failed because she had failed as a parent. Such parents seem to be convinced that placing the blame for learning failure squarely on the shoulders of the school system will miraculously produce a reformed institution providing adequate education relevant to and respectful of their children.

While a consensus is developing that compensatory education is limited in its potential for ameliorating the educational deficits found in disadvantaged children, it is inconceivable that these efforts be abandoned. The social pressure created by the poor, especially in the cities, is simply too great, and social conscience demands further effort. It is essential, then, that we ex-

amine those programs which have had some success in order to find, if possible, key elements that can be employed to increase the potential of compensatory education.

Most compensatory education efforts have focused on curriculum reform, especially on making educational content relevant to the interests of the youngsters. The research project reported in this paper started as a curriculum comparison study; other components, such as program operation and staff model, were held as constant as possible. While the curricula proved to be necessary elements, the outcome of the study strongly suggests that other components are of critical importance. In order to examine this finding in some detail, a description of the project will be presented, followed by a presentation and discussion of the results.

Problem

Since 1962, research projects throughout the country have attempted to determine whether preschool intervention with three- and four-year-olds makes a difference in later school performance. Various projects following several theoretical models have now either reported initial findings or filed final reports (Deutsch, 1968; Klaus and Gray, 1968; Weikart, 1967; Curtis and Berzonsky, 1967; Hodges, McCandless, and Spicker, 1967; Karnes, Teska, and Hodgins, 1969; Di Lorenzo and Salter, 1968). While the cumulative results of these projects offer little to cheer about, the basic conclusion is that the more structured or task-oriented the program, the greater the gains in immediate intellectual competence and, where follow-up data are available, in academic achievement. While further development of programmed curriculum styles and assessment of various intervention methods against "no treatment" control groups are essential, investigation of the relative effectiveness of curriculum models now available is of equal importance. The preschool field has reached a point at which several theoretically divergent curricula may be pulled together in a controlled study to determine their relative impact upon the cognitive, social-emotional, and academic growth of the disadvantaged child.

Although several such comparative studies are underway, little information is available. Karnes (1969) has reported the most extensive data. She found that two specially designed cognitive programs (the Bereiter–Engelmann Language Training project and the Karnes curriculum based on the Illinois test of Psycholinguistic Abilities) were more effective than a traditional nursery school program in promoting intellectual growth. Dickie (1968) reported on the effectiveness of three methods of language instruction in preschool and found no differences between the effects of the unstructured and structured methods; the language training was limited to 20 minutes each day with no carry-over into the 2½ hour total program. Di Lorenzo and Salter (1968) report better success with structured programs of the Bereiter–Engelmann type than with unstructured traditional preschool programs; their study was not designed to explore the impact of differing curricula, however, and this finding is incidental to the total evaluation of their large research

project. Much more information is needed to evaluate the relative impact of various curricula upon the development of disadvantaged children.

Method

The Ypsilanti Preschool Curriculum Demonstration Project was established in the fall of 1967 to document and evaluate three curricula thought to have remedial potential for the disadvantaged. (a) A *unit-based curriculum* emphasizing the social-emotional development goals of the traditional nursery school programs. The hallmarks of this curriculum are introduction of themes and material to acquaint the child with the wider environment, close attention to the individual social and emotional needs of each child, and a considerable degree of permissiveness in classroom operation (Sears and Dowley, 1963). (b) A *cognitively oriented curriculum* developed over the last five years by the Ypsilanti Perry Preschool Project (Weikart, 1967). This is a carefully structured program specifically designed for use with disadvantaged children who are functionally retarded. The curriculum is based on methods of "verbal bombardment," socio-dramatic play, and certain principles derived from Piaget's theory of intellectual development. (c) A *language training curriculum* emphasizing acquisition of academic skills. This task-oriented curriculum, developed by Bereiter and Engelmann (1966) at the University of Illinois, employs many techniques from foreign-language training and includes the teaching of arithmetic and reading. While this program was specifically developed for disadvantaged children, it has not been tried out on functionally retarded youngsters. The demonstration project uses the most recent material published specially for the Bereiter–Engelmann program.

The children for the curriculum study are drawn from the total available population of three- and four-year-old functionally retarded disadvantaged children in the Ypsilanti school district. The contrast group is one of the five no-treatment control groups from the Perry Preschool Project. All treatment groups are balanced by measured intelligence, sex, and race. Two teachers are assigned to each curriculum model after they have had an opportunity to express a preference. They teach class for half a day and then conduct a teaching session in the home of each of their children for 90 minutes every other week. The home teaching phase of the curriculum is executed in the same curriculum style as the classroom program the child attends. Essential to the demonstration aspect of the project is that all three programs have clearly defined week-by-week goals. The curriculum implementation follows a carefully planned daily program designed by the three teams of teachers to achieve the goals of their own curricula. This provision for teacher involvement is a crucial aspect of the overall project.

Results

Results of the project are available from the first year of operation. The replication of the study is now under-

way, and the data from the second year, along with follow-up information on the first year, will be available in the fall of 1969. The data now at hand are based on intelligence-test scores, social-emotional and general developmental ratings by teachers, and systematic classroom observations.

Intelligence tests

Data from intelligence tests indicate the immediate impact of the programs upon the general level of functioning of the children involved. Scores are in no way considered to be indicative of either innate ability or potential capacity. Standardized intelligence tests are easily available indicators of effective programming and, for the population under study here, help to predict later social adjustment in school and academic achievement (Weikart, 1967).

Tables 1 and 2 present the information from the Stanford-Binet, Form LM, as pre-test, and change scores. Both Wave 5[a] and Wave 6 change scores are significantly different across groups when the contrast groups are included in the analysis. However, with the contrast groups removed, no significant

[a] Waves 0, 1, 2, 3, and 4 were part of the original Ypsilanti Perry Preschool Project which started in the fall of 1962. Each curriculum group of Wave 5 (4-year-olds) and Wave 6 (3-year-olds) attended one school together. Wave 5 cognitive-curriculum children had completed two years of preschool at the post-testing. The pre-test scores are from the first Stanford-Binet administration, and the post-test scores are from the most recent testing (June, 1968). The contrast group is the Wave 3 control group of the Perry preschool project taken at ages 3 and 4 and is typical of no-treatment controls throughout the five years of that study.

TABLE 1. Stanford–Binet Scores Wave 5

	Unit (*N*–8)		Cognitive (*N*–11)		Language (*N*–8)		Contrast (*N*–14)		*F*–ratio
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Pre-test	76.4	4.55	75.3	6.06	73.9	5.33	80.8	2.90	
Post-test	94.1	2.42	98.6	12.82	98.2	9.43	84.1	9.70	
Change	17.7		23.3		24.4		3.3		11.3857*

*p .01.

TABLE 2. Stanford–Binet Scores Wave 6

	Unit (*N*–8)		Cognitive (*N*–4)		Language (*N*–8)		Contrast (*N*–14)		*F*–ratio
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Pre-test	73.6	6.93	82.7	5.26	84.4	3.12	80.8	2.90	
Post-test	101.1	7.08	110.7	12.34	114.6	6.14	81.2	10.10	
Change	27.5		28.0		30.2		0.4		25.3940*

*p .01.

differences are found among treatment groups. The three-year-olds in the program of Wave 6 have almost identical I.Q. gains, with a narrow range of 27.5 to 30.2 points gained. The four-year-olds in Wave 5 show a range of 17.6 to 24.4 points gained. Three-year-olds seem to typically record greater gains than four-year-olds, reflecting either the type of items on the Stanford-Binet, the greater malleability of this age group in situations producing change, or the impact of large mental-age changes on a limited chronological age base. The essential point of the table, however, is that all groups gained equally. Indeed, at the three-year-old level almost identical gains were obtained by children in each of the three programs.

Tables 3 and 4 present post-test scores for the three treatment groups on the Leiter International Performance Scale and Peabody Picture Vocabulary Test. Again, all three treatment groups obtained scores that are not significantly different. The results do not show any predictable pattern. The Leiter is in the expected direction for the Wave 5 children. The unit-based and cognitive programs, being the more manipulative and object-experience oriented, have the higher scores; however, this pattern is not repeated with Wave 6 children. The Peabody is clearly in the expected direction for Wave 6, with the Language training children obtaining higher scores; the reverse is true, however, of Wave 5 children. In general, the Stanford-Binet gives the highest estimate of the child's functional level. The essential point is that there are no significant differences in intelligence test scores for the children in the three treatment groups, and the gains are unusually large.

TABLE 3. Wave 5 (4-year-olds) Leiter International Performance Scale and Peabody Picture Vocabulary Test Post-test Scores

	Unit (*N*–7)		Cognitive (*N*–11)		Language (*N*–7)		*F*–ratio	*p*
	M	*SD*	*M*	*SD*	*M*	*SD*		
Leiter	96.0	6.07	93.9	11.10	89.8	9.97	<1	NS
Peabody	94.7	19.65	84.0	20.46	77.3	13.80	<1	NS

TABLE 4. Wave 6 (3-year-olds) Leiter International Performance Scale and Peabody Picture Vocabulary Test Post-test Scores

	Unit (*N*–8)		Cognitive (*N*–4)		Language (*N*–8)		*F*–ratio	*p*
	M	*SD*	*MS*	*SD*	*M*	*SD*		
Leiter	103.2	19.77	112.7	0.83	110.6	7.79	<1	NS
Peabody	78.0	7.61	84.7	9.30	88.4	9.45	1.3656	NS

Ratings by teachers

The teachers in each program were asked to rate all children in their program on two scales: the Pupil Behavior Inventory and the Ypsilanti Rating Scale. These instruments reflect such factors as independence, academic competence, emotional adjustment, and social-emotional state. The ratings were completed by each team of teachers for their own class; as such they are not independent indicators of the actual behavior, etc., of the children. When these data are analyzed according to curricula, there are no significant differences. The children in each of the three programs are seen by their teachers as being much the same in spite of differentiated program focus fostering potentially differentiated modes of adjustment. It is also important to note that the teachers in all three programs rated children who showed academic competence as emotionally adjusted (4 = .67, *p* .01).

Classroom observations

A recent paper by Seifert (1969) reports on systematic observations of classroom behavior in the cognitive and language training programs (the unit-based program was not observed) using the OScAR method. Group teaching sessions were observed on the basis of total statements per minute by teacher, verbal feedback by teacher, amount of pupil-initiated interaction, amount of direct teacher management of pupil, and amount of affect expressed by teacher. The only significant difference found between the two programs was in total statements per minute; apparently the language program does not operate differently, just more intensely.

Thus, in analyzing data from intelligence tests, teacher ratings, and classroom observations, no statistically significant differences are found among the programs. The gains recorded in intelligence tests are unusually high. While long-term follow-up data on school achievement, social adjustment, and eventual disposition of I.Q. level are not available at this time, data from the five-year Perry Preschool Project indicate that children who show early and rapid intellectual growth as a result of preschool intervention also show later social adjustment and academic achievement.

Discussion

These results are unexpected. While there is no special merit in finding no differences among treatment groups, these results, obtained in a compensatory education project with disadvantaged children, raise two critical questions: (1) Why are the changes in intelligence test scores so large; that is, why are I.Q. gains far above those usually reported in the literature? and (2) Why are there no differences in impact among curricula?

It must be stated that these programs really *are* different in their apparent or symptomatic operation. As part of the evaluation program, outside critics representing a wide range of viewpoints are brought to Ypsilanti to appraise the programs. Among these consultants have been E. Kuno Beller, Marion Blank, Courtney Cazden, Joseph Glick, Lawrence Kohlberg, and Todd Risley.

These critics find the programs different in theoretical commitment and differentiated in application. They point out, however, that what is central for one program is often incidental for another. The unit-based and cognitive curricula are more similar to each other in classroom operation than either is to the language training program. The observations of these critics along with the systematic data collected from the approximately 850 visitors who have spent one to five days at the demonstration center in the last year clearly indicate that the programs appear to operate differently. Thus, even though the results of the programs are the same, when the children are measured on general tests, it may be assumed that the operation is actually different for each program.

Why are the change scores so large?

The answer to this question is difficult and would seem to revolve around the selection of the curricula, the staff model for program implementation, and the method of program operation.

1. *Curriculum.* Each of the three classroom units has a clear commitment to a specific theoretical curriculum model. This use of a model provides a framework that sets limits for classroom operation and provides a challenge to the teachers. For example, the cognitive program, derived in part from Piagetian theory, is intellectually challenging and suggestive of specific teaching methodologies. The same is true of the other two curricula to some degree. A framework also helps the teacher select appropriate activities, match program with desired outcome, and fit total classroom operation into a scheme directed at producing specific end products. However, it must again be noted that, at least for the three curricula studied here, the outcomes are the same. The condition of no curriculum, typical of many preschools, has not been tested.

2. *Staff model.* A research and demonstration project produces a fairly specialized environment for staff operation. Since this particular project is aimed at the study of relative curriculum impact, the way the staff operates has been kept uniform in all three programs. Some of the factors directly contributing to the unusually good results are:

a. *Planning.* All teachers must prepare lesson plans at least a week in advance based upon the specific goals of the theoretical framework for their program. These plans have proved to be a daily struggle demanding much in thought and preparation, forcing attention specifically to the use of time in the classroom and to the particular goals for each unit of operation, and providing opportunity for a constant review of curriculum effectiveness. They are available to visitors viewing the demonstration and to supervisors and consultants working with the teachers.

b. *Team teaching.* Two teachers are assigned to each classroom unit for teaching in and operation of one of the program models. Both of them teach all the time and are involved in a continual effort to develop activities and solve problems within the theoreti-

cal framework of the particular model they are employing. This team relationship sharpens the focus on classroom problems by providing an atmosphere of mutual support.

c. *Commitment.* In order to meet the expectations of the project and to be effective in the classroom, the teachers must spend time over and above regular teaching time to stay ahead of the demands. Lunch hours, after-school, "break times," etc., are often used to prepare lessons, write reports, and meet with various staff members and visitors. This type of involvement comes from a firm commitment to the program.

d. *Supervision.* Each team of teachers is supervised by an experienced teacher who works with them to provide focus and to "referee" problems within the team. Rather than smoothing over problems, the supervisor helps the teachers to face the issues and to work out solutions within the theoretical framework of their curriculum model. The supervisor also provides inservice training for the teachers within their curriculum model. While the role of the supervisor is not authoritarian, she is clearly responsible for seeing to it that the teachers keep to the instructional problems at hand.

e. *Respect for individual.* The project is seen as a group of professionals working to produce information. While this group operation ideal has often broken down, the project attempts to keep the staff members in communication. Such interaction has given the staff members an actual part in the development of the total project. It has also kept the project "honest" by forcing all involved to consider all aspects of decisions.

3. *Program operation.* Several things that have characterized the program operation would be expected to have impact on the quality of the results.

a. *Involvement of the mother.* Each of the three curricula includes home teaching as part of the program in order to actively involve the mother in the process of education. While group meetings are held about once a month and some preschool observations are scheduled, the primary focus with parents is the educational activities in the home. The mothers have responded well to these visits and have increased their participation throughout the period of preschool attendance by their youngsters. The staff feels that home teaching has provided powerful supportive action for the children's growth.

b. *Focus on the child.* In order to prepare for the 90-minute home teaching session, the teacher directs her attention to the particular problems of the child before the visit. Upon returning from the home, the teacher writes a report on the visit documenting her observations. The home teaching sessions, therefore, provide an unusual opportunity for the teacher to focus upon the learning problems of each child. This knowledge is carried over into the classroom instructional program.

c. *Focus on education.* The project does not have professional staff other than teachers and research per-

sonnel. It does not offer social-work services, health services, referrals to clinics, agencies, etc. The teachers and the project families see the teacher's role as clearly educational in nature. This single-purpose approach is practical in southeastern Michigan where the services of the many agencies are readily available.

d. *Language.* Essential to the operation of all three curricula is the heavy use of language in the classroom with the students. While the method of language training varies greatly, in all the classes language is used extensively by the adults and is encouraged in the children.

Why are there no differences in impact among curricula?

This question is difficult to answer because the results are much better than expected. Generally, projects of this nature obtain similar results among intervention styles because none of the methods are very effective. This has not been the case with this project. Among the many factors that may have contributed, three seem most crucial:

1. *The staff model and program operation are constant.* In the original design, the curricula were varied, and the staff model and program operation were kept constant. From the initial data collected in this study, it is apparent that the choice of a curriculum framework is only of minor importance as long as one is selected that permits the intensive action suggested by the staff model and the requirements of the program operation.

2. *The curricula are equivalent.* The project data suggest that children may profit intellectually from any structured curriculum that is based on a wide range of experiences. In almost the sense that Chomsky uses in talking about the development of linguistic competence, a child has the potential to develop cognitive skills and good educational habits if he is presented with a situation which requires their expression. Kohlberg (1968) concludes that a child needs broad general forms of active experience for adequate development of his cognitive abilities; he comments that a variety of specific types of stimulation are more or less functionally equivalent for cognitive development. These three curricula, as diverse as they seem to be, apparently are equivalent.

3. *The staff expectations for the children are uniformly high.* Much has been said recently about the "Rosenthal effect"* and the impact of motivational changes on preschool outcomes when assessed by standard tests. Certainly a portion of these gains has been produced by these factors. For example, Zigler (1968) identified a change of about six to ten I.Q. points as a product of improved motivation. Rosenthal has reported impressive gains in test performance by children in early grades labeled as "bright" for teachers by outside researchers. While these factors contributed to the size of the gains reported here, it is assumed that they were operating equally in all programs

*For a critique of Rosenthal's position see Snow, R. F. Unfinished pygmalion, *Contemporary Psychology,* 1969, 13, 197–199.

and that such gains were only a portion of the total.

Conclusions

The basic implication of these findings is that a shift in focus is necessary for both preschool education and compensatory education. The heavy emphasis on curriculum development, while important, has greatly overshadowed the need for careful attention to both the staff model and the program operation. Both the mechanical application of a specific curriculum and the busy concern with administrative procedure that any program operation entails can doom a project to failure.

For preschool operation, these findings mean that a staff is free to develop or employ any active curriculum that it believes will match the needs of the children so long as that curriculum provides adequately for staff involvement and facilitates the type of program operation desired. The arguments about the relative effectiveness of the various approaches to preschool education are irrelevant. Then too, waiting for *the* curriculum for disadvantaged children to be developed so that early education programs can be effective is pointless. The process of creating and the creative application of *a* curriculum, not the particular curriculum selected or developed, is what is essential to success.

In addition, program operation must include careful attention to three areas. First, the program must include opportunity for the teacher to intensively think about each child in the project. Teachers apparently treat the educational development of young children more effectively if they evolve an intimate knowledge of how a child learns and responds through their own direct experience with that child. Second, the project must provide a way to include mothers in the educational process. This is not so much a transfer of information or experience to the mother as an attempt to create an atmosphere of support for intellectual growth in the home. Third, the staff model employed must allow opportunity for each individual to be creatively involved in the total operation. While administrative direction and a good curriculum are important to achieving success, staff involvement is crucial. In an almost romantic sense, the human involvement of concerned teachers and staff is the key element in program success.

Featherstone (1969), in a recent article, comments on the British Infant Schools:

> *But the danger I'm most anxious to avoid is leaving the impression that one can single out a few elements of a good school and turn them into a formula to impose on teachers and children in other schools. There is no single lever to pull, or technical solution. What we can do is work toward an idea of the kind of learning we wish to promote. That, among other things, is a matter of choosing what we value.*

Our data agree with Featherstone's observation that there is no single lever to pull and certainly no technical solution. Compensatory education can reach the child through a range of pro-

grams appropriate to him. To be effective, however, it is necessary that the programs be organized and operated in such a manner as to allow the full utilization of human insight and commitment.

Addendum

Since this paper was delivered, research comparing the three methods has continued. According to a personal communication received from the author, data gathered over a two-year period show that the performance of children in the Unit program tends to lag behind those in the Cognitive and Language programs.* However, the author attributes this disparity in results not so much to essential differences among the curricula as to difficulties in maintaining early momentum and enthusiasm in the Unit program. Teacher behaviors tends to be fairly specific in the Cognitive and Language programs, and it is easier to supervise teachers and help them maintain performance levels. The Unit program, on the other hand, is highly permissive and loosely structured, and it is difficult for supervisors to determine whether teachers are "following the program" in time to offer corrective assistance. In the author's opinion, the failure to provide adequate supervision within any focused curriculum framework is the major reason why Head Start and similar programs have failed in many communities.

*This effect began to show up in Wave 6; see Tables 2 and 4.

References

Bereiter, C. and Engelmann, S. *Teaching Disadvantaged Children in the Preschool.* Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Curtis, C. A. and Berzonsky, M. D. Preschool and primary education project. Council for Human Services Commonwealth of Pennsylvania, Harrisburg, Pa., 1967.

Deutsch, M. Interim progress report, part II. Institute for Developmental Studies, New York University, New York, 1968.

Dickie, J. P. Effectiveness of structured and unstructured (traditional) methods of language training. In Brottman, M. A. (Ed.) Language remediation for the disadvantaged preschool child. *Monographs of the Society for Research in Child Development.* No. 124, 1968, 62–79.

Di Lorenzo, L. T. and Salter, R. An evaluative study of prekindergarten programs for educationally disadvantaged children. *Exceptional Children,* 1968, **35**, 111–119.

Featherstone, J. Why so few good schools? *New Republic,* January 4, 1969, 18–21.

Hawkridge, D. G., Chalupsky, A. B., and Roberts, A. O. A study of selected exemplary programs for the education of disadvantaged children. Office of Education, United States Department of Health, Education, and Welfare, 1968.

Hodges, W. L., McCandless, B. R., and Spicker, H. H. The development and evaluation of a diagnostically based curriculum for preschool psychosocially deprived children. United States Department of Health, Education, and Welfare, 1967.

Jencks, C. An alternative to endless school crisis—private schools for black children. *The New York Times Magazine,* November 3, 1968.

Karnes, M. B., Teska, J. A., and Hodgins, A. S. A longitudinal study of disadvantaged children who participated in three different preschool programs. Paper presented at the annual meeting of the American Edu-

cational Research Association, Los Angeles, California, 1969.

Klaus, R. A. and Gray, S. W. The early training project for disadvantaged children: a report after five years. *Monographs of the Society for Research in Child Development,* No. 120, 1968.

Kohlberg, L. Early education: a cognitive-developmental view. *Child Development,* 1968, 39, 1013–1062.

Sears, P. S. and Dowley, E. M. Research on teaching in the nursery school. In Gage, N. S. (Ed.) *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963, 811–864.

Seifert, K. A comparison of verbal interaction in two preschool programs. Unpublished manuscript, University of Michigan, 1969.

Weikart, D. P. (Ed.) *Preschool Intervention: Preliminary Report of the Perry Preschool Project.* Ann Arbor, Michigan: Campus Publishers, 1967.

Zigler, E. and Butterfield, E. C. Motivational aspects of change in I.Q. test performance of culturally deprived nursery school children. *Child Development,* 1968, **39**, 1–14.

4.6 *Modern Learning Theory and the Elementary School Curriculum*

PATRICK SUPPES

Reprinted from the *American Educational Research Journal,* 1964, **1,** 79–93, with permission of the author and the American Educational Research Association. Patrick Suppes is an educational psychologist who directs the Institute for Mathematical Studies in the Social Sciences at Stanford University. His special interests are concept formation in children, mathematical models in social interaction, and the philosophy of science.

The substance of this paper was given as an invited address to the annual American Educational Research Association meeting in Chicago in February, 1964. It describes some of the approaches to applying learning theory via computer technology to classroom instruction in what has since been called "computer-assisted instruction," or CAI.

Introduction

What I want to do this evening is to sketch for you some of the implications of recent research in learning theory for the elementary-school curriculum and to indicate how I think these implications can best be realized by appropriate application of modern computer technology. I shall say a number of things about computers, but I want to emphasize at the very beginning that I do not intend to let the fascinating problems of computer technology distort the primary focus on learning and the curriculum. The research and curriculum work done thus far as a result of the current interest in teaching machines and programed instruction show clearly that technology alone is not going to produce any fundamental or long-lasting changes in the curriculum. Only if its applications are guided by appropriate psychological principles and subject-matter insight will such changes be brought about.

To show that I mean what I say about keeping technology at heel, I will turn directly to certain results derived from recent work in learning theory and indicate some of their pedagogical implications. Only after that has been done, will I discuss how computers can be used to implement these results.

Individual Differences in Learning

One of the most firmly established generalizations of behavioral psychology is the existence of significant differences in individual rates of learning. In spite of the obeisance paid to this tenet in discussions of curriculum, I consider it the most important principle of learning as yet *unaccepted* in the day-to-day practice of subject-matter teaching in the classroom. In many large elementary schools, homogeneous groupings of children at the same grade level is done routinely. The recent movement toward the ungraded elementary school constitutes a significant development for the accommodation of individual differences in learning. However, I venture to suggest that the staggering implications of these differences have scarcely been realized in any school in the country.

Let me give an example from some recent pedagogical experience of my own. Beginning with the current academic year we have undertaken at Stanford to give an intensive accelerated program in mathematics to a group of gifted first graders for the 6 years of their elementary-school attendance. The initial group of 40 students was selected by giving the *New York Test of Arithmetical Meanings* to all the children entering the first grade in four schools. Those above the 70th percentile on this test of first-grade achievement were then tested more intensively, most of them being given a form of the Stanford-Binet. The 40 children selected by this procedure have IQ's ranging from 122 to 166 with a mean of 137. They are being taught mathematics in the four schools in small groups of 9 to 12 pupils each. For each child, we are collecting subject-matter learning data, consisting of a daily list of the problems worked and the problems answered incorrectly. Those of you who have attempted

detailed behavioral analysis of subject-matter learning will know that this is no small task, particularly when the pupils are bright and therefore able to move rapidly through the course work. Because the children are being handled in small groups, it is possible to give them something close to individual treatment. From our standpoint, the most significant aspect of this individual treatment is the fantastic differences in rate of learning. At the end of the first four weeks, the fastest child in the group covered approximately half again as much material as the slowest (and I hasten to add that the relative rates of progress are not strongly correlated with the IQ scores). To give you a sense of numbers, during the first 7 weeks (the classes meet 4 days a week with each session lasting not more than 35 minutes) the fastest child worked approximately 3,400 problems with an error rate less than 2 per cent. During this same period the slowest child in this highly selected group worked approximately 2,200 problems. Learning curves for these two students are shown in Figure 1. The number of weeks of participation is shown on the abscissa and the cumulative number of problems worked on the ordinate. A calibration of the number of problems in terms of the *Sets and Numbers* elementary-school mathematics series I have been writing is also shown on the ordinate. Books 1A and 1B constitute a full first-grade curriculum, and Book 2A the first half of a full second-grade curriculum. At the end of the seventh week, these two bright first graders are already separated by almost a third of the second-grade curriculum. For understandable

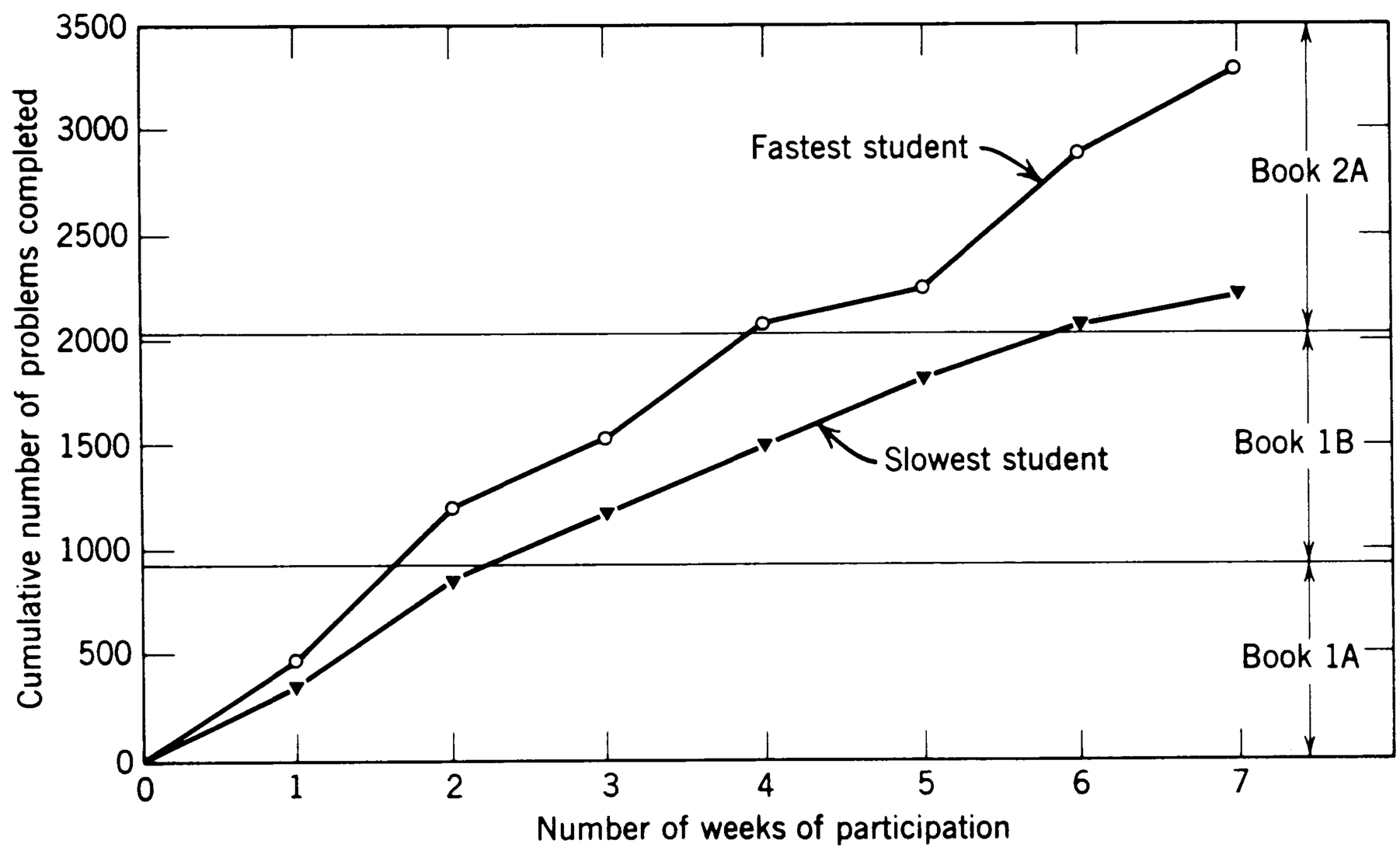


FIGURE 1. Progress of fastest and slowest of forty gifted first-grade pupils in mathematics.

reasons, we are reluctant to forecast the maximum separation after this class has continued for another couple of years. The rate at which these bright children are working problems and proceeding through the curriculum is most surprising; it has certainly exceeded our initial expectations. We are not encouraging a speed contest; in fact, the teachers are making a decided effort to discourage intensive competition. These comparisons are, I think, a fair representation of the very large differences to be expected among even very bright children when the rate of progress of the individual child is not tied to that of the group.

Because we wanted to know what learning rates to anticipate, we attempted to search the literature on mathematics learning in the elementary school. It is disappointing to find how few hard data are reported on the relative rates of learning of bright or slow children under teaching conditions approximating individual treatment.

To underscore what is probably obvious to all of you anyway, let me describe a recent experiment I performed jointly with Edward J. Crothers on the beginning stages of learning to read by kindergarten children. In contrast to the mathematics project just described, this was a controlled experiment with all subjects being run by an individual investigator in an isolated experimental room. The subjects were not selected for ability, but were all kindergartners in an elementary school near Stanford. To build up a working vocabulary without requiring many months of training, we selected 14 letters of the alphabet from which a reasonable number of familiar words could be constructed. I won't attempt to describe all aspects of the experiment, which in total design was fairly complicated. For the purpose of my present remarks the essential facts are these. The subjects were trained to a criterion on 14 single words built up from the 14 letters, the criterion being 28 successively correct responses. After this criterion was satisfied, they were given a list of 14 two-word phrases, then three-word phrases, then four-word phrases and five-word phrases, all built up from the initial vocabulary of 14 words. To obtain quick and accurate error data, the subjects were trained in the following way: on each trial the experimenter said a word or phrase and the child responded by pointing to one of three possible graphemic representations of the auditory stimulus. (I emphasize that this experimental procedure is not recommended as a method of teaching reading. It was an efficient and objective procedure for studying the acquisition of the initial phases.)

Thirty-eight kindergarten children completed the experiment. The point I want to emphasize is the wide range in the total number of trials required. The fastest child needed only 196 learning trials while the slowest child needed 2,506. This range is not misleading, for the mean was 967.4 with a standard deviation greater than 400.

Data of this sort, which may be multiplied tenfold by references to the literature, argue that by far the greatest improvement in subject-matter learning will result from an almost single-minded concentration on individual dif-

ferences. I am sure that the optimum sequencing of curriculum materials, the analysis of the subject matter so as to present it in steps of the proper size, etc., are not nearly so important elements in learning as the single one of individual accommodation. I shall have more to say on this matter when I discuss computer technology.

Immediate Reinforcement and Overt Correction Procedures

A variety of experimental studies has shown the importance of both immediate reinforcement and overt correction in children's learning. As a typical example, let me mention one experiment we conducted three years ago at Stanford (Suppes and Ginsberg, 1962). There are good studies (e.g., Burke, Estes, and Hellyer, 1954) to show that requiring an adult subject to make an overt correction response after reinforcement neither increases his learning rate nor influences appreciably his asymptotic behavior. Our problem was to determine whether these findings would hold for primary-grade children. The experiment concerned recognizing the abstract patterns involved in notation for any number system—in this case the numbers four and five in binary notation. On each trial a stimulus was presented to the subject to which he could make one of two responses: he could place the stimulus on top of the Arabic numeral 4 or on top of an Arabic numeral 5. Six different stimuli were presented, three of them (*e.g.*, $\alpha\beta\beta$, $\lambda\pi\pi$, $\delta\varepsilon\varepsilon$) representing four in the binary number system and the remaining three (*e.g.*, $\alpha\beta\alpha$, $\lambda\pi\lambda$, $\delta\varepsilon\delta$) representing five. There were 48 subjects in the study, half of them coming from grade one and half from a kindergarten class. Twenty-four of the subjects, 12 kindergartners and 12 first graders, were asked to make an overt correction response after a wrong choice, that is, after placing a stimulus card on the wrong Arabic numeral. The other 24 subjects were not required to make such an overt correction response, although they were told on each trial whether or not their responses were correct. The learning curves for the two groups showed a significantly faster rate of learning for the overt correction group throughout the entire experiment, which consisted of 96 trials. A *t*-test computed on the over-all responses of the two groups was significant at the .001 level.

Other studies demonstrating the desirability of overt correction procedures to facilitate children's learning may easily be cited. On the other hand, those of you who have done much observing of the environment of elementary-school classrooms, realize fully the difficulties facing the teacher in attempting to provide immediate reinforcement and overt correction. If we take a typical lesson in arithmetic, a little simple arithmetic itself will show these difficulties. As a conservative estimate, it is not uncommon in the latter part of the first grade for a workbook page to contain ten problems. Let us suppose, again conservatively, that the class consists of only 25 children. If the class is working this particular page together, the teacher cannot hope to monitor the 250 problems being done and, on the spot, have each child

make appropriate corrections. Some of the elementary-school mathematics teachers with whom I have worked have told me that in order to do a good job in the first or second grade it is necessary to spend at least one hour after each school day correcting the children's workbooks. Some of them insisted that a more realistic estimate would be two hours. I have watched teachers who have gone through this laborious process of marking each child's work on each problem and requiring him to correct his errors the next day. I wish I were able to report some experimental data on the efficacy of overt correction delayed by at least one day. The impression of the teachers who have tried this procedure has been positive, but I cannot cite any controlled data. Again, however, I want to emphasize that technological methods are almost necessary in order to provide overt correction procedures, particularly with immediate reinforcement, on a sustained basis. This is all the more true if we do not restrict ourselves to mathematics but think about the other parts of the curriculum, such as foreign language, elementary science, and social studies.

Specific Nature of Transfer

One of the most important and fundamental measures of learning is the extent of transfer to new situations. The classical paradigm for evaluating the benefits of a new curriculum is in terms of transfer, with the experimental group receiving training of Type A, the control group training of Type B, and both groups then being tested on transfer items of Type C. A comparison of performance on the transfer items is the primary method of judging the efficacy of the new treatment given the experimental group.

Transfer and generalization have been a major focus of investigations by learning theorists. In almost all the fundamental work on stimulus generalization and transfer in the literature of mathematical learning theory during the past ten years (e.g., that by Estes, Bush and Mosteller, Restle), the similarity between stimulus patterns, or displays, is represented by the amount of overlap between two sets of stimulus elements. In the simplest experimental paradigm, only two stimulus situations are treated. Training takes place on one, and generalization or transfer to the other is tested. The amount of generalization or transfer is taken to be a measure of the amount of stimulus overlap in terms of the set of stimulus elements common to the two situations. A number of ingenious and important experiments of this general type have been conducted within the domain of learning theory in the past fifteen years, but it is not appropriate to review them here.

There is, however, a common failure in the reports of these experiments, which is reflected in the paradigm of educational research sketched above. This common failure may be characterized as a lack of structural depth in theory. It is a well-known mathematical fact that sets without structure provide little basis for the elaboration of a theory.

Experiments on mathematical concept formation provide an excellent op-

portunity to dig deeper into the structure of the *set* of stimuli, because in most instances of such concept formation a natural and clearly defined structure is imposed by the mathematical character of the concept itself. Let me give an example that I have studied experimentally in collaboration with Rose Ginsberg (Suppes and Ginsberg, 1963). Given any pair of sets or classes of objects, we may classify the pair of sets in one of the four following mutually exclusive and exhaustive ways. The first possibility is that the two sets are identical in terms of their members and in terms of the order assigned to these members. The second possibility is that two sets are identical in the sense of having the same members, but the members are not given in the same order. For example, the set consisting of the numbers 1 and 2 is identical in this sense to the set consisting of the numbers 2 and 1. The third possibility is that the two sets are *equivalent* but not identical; that is, they have the same number of elements but the elements are different. For example, the set consisting of 1 and 2 is equivalent but not identical to the set consisting of 3 and 4. The fourth and final possibility is that the two sets are not equivalent; that is, they do not have the same number of elements. These four possibilities, or categories, may be used as four subconcepts for analyzing the learning of more general concepts.

In one of our experiments concerned with these subconcepts, first-grade subjects were given 56 trials on identity of ordered sets and then 56 trials on identity of sets without concern as to the order of the members. Similar subjects were given 56 initial trials on identity of unordered sets. The sets pictured by the stimulus displays consisted of one, two, or three elements. On each trial, two of these sets were displayed. Subjects were told to press one of two buttons—one button when the two stimulus displays presented were "the same" and the other button when they were "not the same." When the correct response was made, the appropriate one of two reinforcing lights flashed. Using simple principles of interference, we might expect negative transfer in going from identity of ordered sets to identity of sets. That this rather unsurprising result did occur is easily seen by comparing the mean learning curve for the children who began the experiment on identity of sets with the curve for the children who learned identity of sets *after* identity of ordered sets. The evidence of negative transfer from comparison of these learning curves is clear. The interesting question of greater depth is: can we identify the point of negative transfer in terms of the structure of the four subconcepts described above? The answer is clearly affirmative. In terms of the four categories, the pairs that are identical but not in the sense of order and *only these* require a different response in passing from the subconcept of identity of ordered sets to that of identity of unordered sets. And all the negative transfer is indeed isolated in the pairs that are instances of this subconcept. For the other three subconcepts—identity of ordered sets, equivalence of sets, and nonequiva-

lence of sets—there is, in fact, slight evidence of positive transfer. Detailed data are presented in Suppes (1963b). The highly specific nature of the transfer results obtained here is undoubtedly present in other concept learning.

The identification of the structure of subconcepts determining the nature of transfer is a central problem for our future work in mathematical concept formation, and I hope soon to have more results to report.

Optimal Block Size in Learning a List of Simple Items

The learning task I want to take up under this heading is, on the face of it, a very simple one. We want subjects to learn a list of items. The list is rather long, so we may consider the possibility of breaking up the list into blocks of varying size.

A number of applications of the results of studies of this kind can be mentioned. Among the most interesting are those arising in learning a second language. The items may be foreign words, and the task is to learn their approximate meanings in English. In psychological terms, this means that to each word of the foreign language in the list the subject is asked to associate a unique English word. We all know that the words and phrases of a second language vary in difficulty. In graphemic representation some words are almost identical in the two languages although at the phoneme level they may differ considerably; e.g., "application" is graphemically the same but phonetically different in English and French.

In the present discussion, I want to ignore all these differences in difficulty. Suppose we want subjects to learn an initial Russian vocabulary of approximately 100 words. How should we work our way through this list of words in order to maximize learning? We may fix the total number of trials, with the terminal trials used as test items. The problem is to decide how the remaining trials, apart from the test, should be allocated for training purposes. At the end of the training trials, each item will have a mean error probability and in order to maximize learning we want to minimize this mean error. Items are to be given in blocks of a fixed size, and our problem is to determine what this size should be.

Fortunately, under quite general assumptions about learning and forgetting, some recommendations about block size can readily be reached. Assuming that learning is faster than forgetting, we can show that the block size should be as large as possible, which means that in the case of our 100 Russian words the subjects should be taken through the entire list once before a second presentation of any item. On the other hand, when learning is slower than forgetting, the block size should be as small as possible; that is, all exposures of a given item should be contiguous. In most situations, learning seems to be faster than forgetting and thus, according to the theory, the largest possible block size should be used.

The formal basis of these results is

easily described in mathematical terms. Let $p_{i,n}$ be the probability of a correct response to item i on trial n, and let $q_{i,n}$ be the probability of an incorrect response to item i on trial n. Then

$$q_{i,n} = 1 - p_{i,n}.$$

When an item is presented and the correct response reinforced,

$$q_{i,n+1} = \alpha q_{i,n};$$

i.e., there is learning and thus a decrease in the error probability. (We may interpret $q_{i,n}$ as the *average* probability of an error if we wish to hold an all-or-none view of the learning of individual items by individual subjects.)

On trials when other items are presented, there is forgetting and thus a decrease in $p_{i,n}$:

$$p_{i,n+1} = \beta p_{i,n}.$$

The learning parameter is α and the forgetting parameter is β, and $0 \leq \alpha$, $\rho \leq 1$. When $\beta > \alpha$, learning is faster than forgetting; learning is fastest when $\alpha = 0$, and therefore the largest possible block size should be used. Mathematical details of this derivation are given in Suppes (1963a). From a psychological standpoint, an analysis of this sort is closely related to the classical experimental literature on part vs. whole learning. See, for example, Woodworth and Schlosberg (1954).

When I first obtained these simple results about block size based on a purely theoretical argument, I was skeptical that they would hold for a very wide range of learning phenomena. However, Edward J. Crothers, Ruth Weir, and I have recently been testing them in some experimental work with Russian and they have held up well. In the first experiment completed, the task presented to subjects was learning 108 Russian words. On each trial the subject heard the Russian word spoken in a short phrase. At the same time he was shown three written translations of the phrase, with the word in question being the source of variation. Four groups of subjects were run. One group received the vocabulary items in blocks of 6, the second group in blocks of 18, the third group in blocks of 36, and the fourth group in blocks of 108 (that is, the total list). Each item was given 21 reinforcements and thus there were about 2,200 trials for each group. It required 13 days of one-hour sessions for subjects to complete the experiment. The group of subjects that had blocks of 108 performed slightly better than the other groups. As expected theoretically when learning is faster than forgetting, the 6-block group was the worst.

We are now testing the theory with a second experiment using a basic vocabulary of 216 words. This experiment is not yet complete, but at present the 216-block group is doing better than the 108-block group.

This work on Russian has been conducted with college and junior-high-school students. We are now planning some work on the learning of Russian by elementary-school pupils, and our planning of this program has been very much influenced by the experiments done thus far with older subjects. I expect our main findings to hold for this younger group as well.

Response Latency as Criterion of Learning

Although interest in response latencies, or reaction times, is a very old one in psychology, it is surprising how little systematic work has been done during the past decade in this area, either on a theoretical level or in experimental studies. Only during the last year or two has there been a serious revival of interest in response latency as a variable that may serve as a central criterion of learning. Some of this recent experimentation is showing clearly what all of us probably know in a muddle-headed way: namely, that, for many kinds of learning, latencies are a much more sensitive index than are response errors themselves. For at least three areas of the elementary-school curriculum—reading, mathematics, and foreign language, there is no doubt that response latencies are more sensitive measures of skill mastery and depth of learning than the responses themselves. A certain amount of work on latencies has been done with respect to reading, but not of the sort that is very satisfactory from an experimental standpoint. In the case of elementary mathematics, I have recently had an assistant searching the literature. He has found very little that is directly relevant. Since what happens to response latencies under continued practice of the algorithms of arithmetic is such a natural subject for investigation, I can hardly believe that detailed studies in this area do not exist. I have been particularly surprised not to find a substantial body of literature from the twenties on this topic. In any case, an extensive series of studies beginning with first-grade children is now under way in our laboratories, and by this time next year I may be able to report some interesting results. Interest in response latencies in the doing of simple arithmetical calculations is not a matter of sheer empiricism. A few elementary computations show clearly that even the facts of arithmetic involved in sums and differences of numbers not greater than 10 are far too numerous to be simply stored in memory. Simple algorithms of some sort are almost surely learned in working problems like "4 plus what number is equal to 7?" Structural hypotheses about the nature of these algorithms are not easily verified. One of the few possibilities for getting evidence in this area lies in the quantitative analysis of latency data. It is fair to claim that as yet we know very little about how algorithms are learned, and until a fundamental theory of some validity is developed, we shall continue to stumble around in much of our discussion of how arithmetic should be taught.

There have been a few isolated studies of response latencies in second-language learning; for example, Lambert (1955) has done some interesting work on the measurement of "fluency" in bilinguals. However, in view of the obvious relevance of response latencies to measures of second-language mastery, the paucity of serious experimental studies in the literature is almost shocking.

To indicate some of the ways in which response latencies are sensitive

measures of learning, let me describe a recent study of ours dealing with the recognition of spoken Russian words. On each trial the stimuli were spoken by a native speaker. The Cyrillic response alternatives were shown in multiple-choice format. When the subject heard a word, he pressed the key corresponding to one of the alternatives and was immediately informed of the correct answer. As an independent variable, the Russian items were classified according to the similarity of their constituent printed characters to one another and to the letters of the Roman alphabet. For example, we put in one category Russian words whose graphemic representation is composed entirely of letters that also occur in the Roman alphabet. On every trial the response latency as well as the response of the subject was recorded. We found that latencies strongly depended on the type of item. For example, longer latencies were associated with items having Cyrillic characters similar to each other and distinct from Roman letters. It was possible to divide the items into five categories and have for all sessions a clear separation of the latencies. Most important, even after the response-latency curve had flattened out to an asymptotic level by essentially the tenth session, the clear separation in response latencies for the five types of items remained. A foremost problem for language teaching is to devise methods for breaking through these asymptotic levels of response latency by finding methods that will, with sufficient practice, drive them down to the asymptotic latency levels of the subject's native language.

Computer Technology

I said that I would have a fair amount to say about computer technology and the possibilities it opens up for application of learning theory to the elementary-school curriculum. Let me begin by describing the computer-based laboratory for learning and teaching we are now constructing at Stanford. (The Executive Committee of the laboratory consists of Richard C. Atkinson, William K. Estes, and me.) Then I shall indicate how the laboratory may be used to implement some of the research findings already discussed.

The core of the laboratory is a small fast computer, which controls a variety of terminal equipment located in six student booths. Because the computer is used in a time-sharing mode, each student works independently of the others. He faces a 16-inch cathode tube that looks like a television screen. This cathode tube—"scope" as we call it—is connected directly to the computer, and visual stimulus displays are generated directly from programs in the computer. One of the great advantages of the scope is that the student's responses to a stimulus display may be immediately made an integral part of the display itself. The student has two methods of response available: one is to use a light probe to make selections or indicate choices on the face of the scope; the other is to use a standard keyboard that is placed just below the scope.

The second important device for the visual display of stimuli is a microfilm unit. The equivalent of a 512-page book may be encoded on microfilm and any

quarter of a page may be reached under random access within one second for display on a visual frame slightly larger than the standard 8½″ by 11″ page. The student may respond to a frame presenting a problem by using a light probe on the face of the display itself. Multiple-choice responses are not the only kind available with the microfilm equipment; the possibilities of constructed responses are quite substantial. As part of the visual display of a page, we may show a keyboard at the bottom of the page and the student may respond on the keyboard by indicating the appropriate sequence of letters with his light probe.

Since visual stimuli are not sufficient for young students, we can also present auditory stimuli—either through individual speakers in each student's booth or through high-fidelity earphones. The auditory system will be of sufficient fidelity to permit work with foreign languages. Finally, we plan to have a closed-circuit-television setup, so that we will be able to show the students a televised instructor. I emphasize that both the auditory stimuli and the instructional material on videotape will be under computer control, which means that the material will not have to be worked through in a simple linear order, but can be presented according to the needs and learning rate of the individual student.

What we hope to obtain with this computer-based laboratory is a reasonably good simulation of the teaching environment created by a tutor working with an individual student. In terms of the five phases of learning research I have mentioned, let us see how this computerized environment will facilitate presentation, organization, and behavioral analysis of curriculum material.

The application to individual differences, which I mentioned first, is apparent. Since each student will proceed at his own pace (independent of the progress of the other students), individual differences can easily be accommodated. The computer-based terminal equipment can also provide immediate reinforcement and overt correction when necessary. In different parts of the curriculum, different approaches to immediate reinforcement and overt correction may be adopted, but it should be clear how easily and directly a variety of procedures may be instituted. Let me give just one example, which concerns the teaching of place value in the first grade. The child is asked to decompose a three-digit number into the number of hundreds, the number of tens, and the number of ones. He has been shown on the scope the sentence,

456 is____hundreds____tens____ones.

Above the hundreds place is a small star indicating, like the bouncing ball of singing movies and television, that this is the first blank he is to fill. If he makes the correct response, the numeral 4 is immediately placed in the blank and the star then moves to the next blank. If in making the next response he makes an error, a large X appears and he knows that he must continue to respond until the correct number is given.

For the past several years we have been conducting an experimental pedagogical program in the teaching of

mathematical logic to able fifth and sixth graders. We have just recently begun to experiment with how this program may be adapted to the special environment of our new laboratory. Let me describe how we can apply immediate reinforcement and correction even to the variety of individualized mathematical proofs students construct very early in their training. Also I would like to indicate how we have begun to use the computer to reduce, rather than increase, the amount of tedious work on the part of the student. A good portion of the course consists of formally deriving conclusions from given premises. When the premises are symbolized in mathematical form, the student in the usual teaching situation is asked to write out each line of derivation and to justify it according to an appropriate logical rule of inference. With the facilities available in the new laboratory we have been able to shorten this procedure and to require the student to think only of which rule of inference to apply and which lines he wants to use to get the new line of derivation. He then types in on the keyboard the abbreviation for the rule and the numbers of the lines he is going to use. The computer does the rest as part of its teaching program and displays on the scope the new line of proof. If an incorrect response is made, the reason why the rule typed in is inappropriate is displayed in place of the line requested by the student. I emphasize that in general the correct response to be made by the student is not unique. Any appropriate response is accepted, i.e., any response that satisfies the criterion of being a logically valid application of the rules of inference. It is most impressive to watch the speed with which able fifth graders accommodate to this new environment and rapidly proceed to work a large volume of problems.

In relation to my earlier remarks about transfer, the most immediate application of the new laboratory is to permit an efficient analysis of transfer among concepts over a substantial period of time. If we want to follow 30 students, let us say, through the first two years of the elementary mathematics curriculum, the task of collecting, organizing, and analyzing response data is almost overwhelming without a computerized facility. One of our first aims will be to obtain this sort of data in order to make a deeper analysis of transfer phenomena from one mathematical concept to another as the child progresses through the early phases of his mathematics education. We are particularly interested in seeing what the effects will be of weak and strong criteria of attainment on each concept before passing to another concept. How will the variation in criterion of performance affect transfer to a new concept and subsequent learning?

The application of results of optimization to the organization of stimulus material scarcely needs detailed comments. In students' work with the elementary parts of a foreign language, for example, it is apparent that a properly prepared computer program can be sensitive to points of difficulty in a student's performance, certainly at the level of language recognition, in at least as effective a way as all but the

best language teachers. The major criticism I would make, from a learning standpoint, of the vast majority of language laboratories is this. The behavioral aspect of language learning has not been made an integral part of the work. Too often the student is *passively* listening to linguistic material, when it is a simple matter with any sort of computer facility to arrange for recognition responses to be made and evaluated. In learning to recognize spoken words and phrases in a foreign language, the student can be given repeated opportunities to make the appropriate identification. Those items he finds easy can be dropped from the list of items presented and new items inserted as the occasion demands. Even the best teacher would have difficulty keeping up with the rapid tabulation and decision to insert new items that can be made with a computer program.

The same comparison in favor of the computer-based laboratory can be made when we attempt to apply response latency as a criterion of learning. Even the best individual tutor working with an individual student finds it difficult to make an accurate appraisal of response times. An approximation that is not too bad can be made in the case of arithmetical skills—for example, speed skills in doing the longer algorithms of arithmetic—because the times involved are relatively long. The situation is quite different in terms of language recognition and production. The response times involved are short, on the order of one second, and it is not an easy matter for the experienced tutor to make an accurate discrimination about the performance of the learner. I daresay that, in the latency experiment I described, none of us here would have made a very successful discrimination in the response times for the five different kinds of items in terms of their relation to English sounds and graphemes. Moreover, the imposition of a sophisticated latency criterion of performance can be made with great ease in a computer-based laboratory but is extremely difficult even in work with an individual child by an experienced tutor. It is our own conjecture that the use of latency criteria will be one of the most important new aspects in which we shall be able to deepen our understanding of the relation between learning and the curriculum.

I have emphasized various detailed ways in which the computer-based laboratory can facilitate the applications of learning theory to the school curriculum. Let me conclude by mentioning the over-all point of greatest importance for the immediate future. A computer-based laboratory such as I have described is one of the few ways of building a satisfactory bridge between research in learning and curriculum work. The reasons are really not scientific, but technological. The difficulties of collecting an adequate amount of behavioral data on subject-matter learning are so great and the problem is so complex that it is difficult to conceive of doing an adequate job with simpler apparatus. The existence of large bodies of organized data will surely promote a more intimate interaction between systematic theories of learning and their applica-

tion to education. The relation of psychology to education will come closer to what Dewey forecast it should be many years ago (McLellan and Dewey, 1895), "In a word, education itself is precisely the work of supplying the conditions which will enable the psychical functions, as they successively arise, to mature and pass into higher functions in the freest and fullest manner, and this result can be secured only by knowledge of the process—that is, only by a knowledge of psychology."

References

Burke, Cletus J.; Estes, William K.; and Hellyer, Sidney. "Rate of Verbal Conditioning in Relation to Stimulus Variability." *Journal of Experimental Psychology,* **48**: 153–61; September 1954.

Lambert, Wallace E. "Measurement of the Linguistic Dominance of Bilinguals." *Journal of Abnormal and Social Psychology,* **50**: 197–200; March 1955.

McLellan, James A., and Dewey, John. *The Psychology of Number and Its Applications to Methods of Teaching Arithmetic.* New York: D. Appleton, 1895, 309 pp.

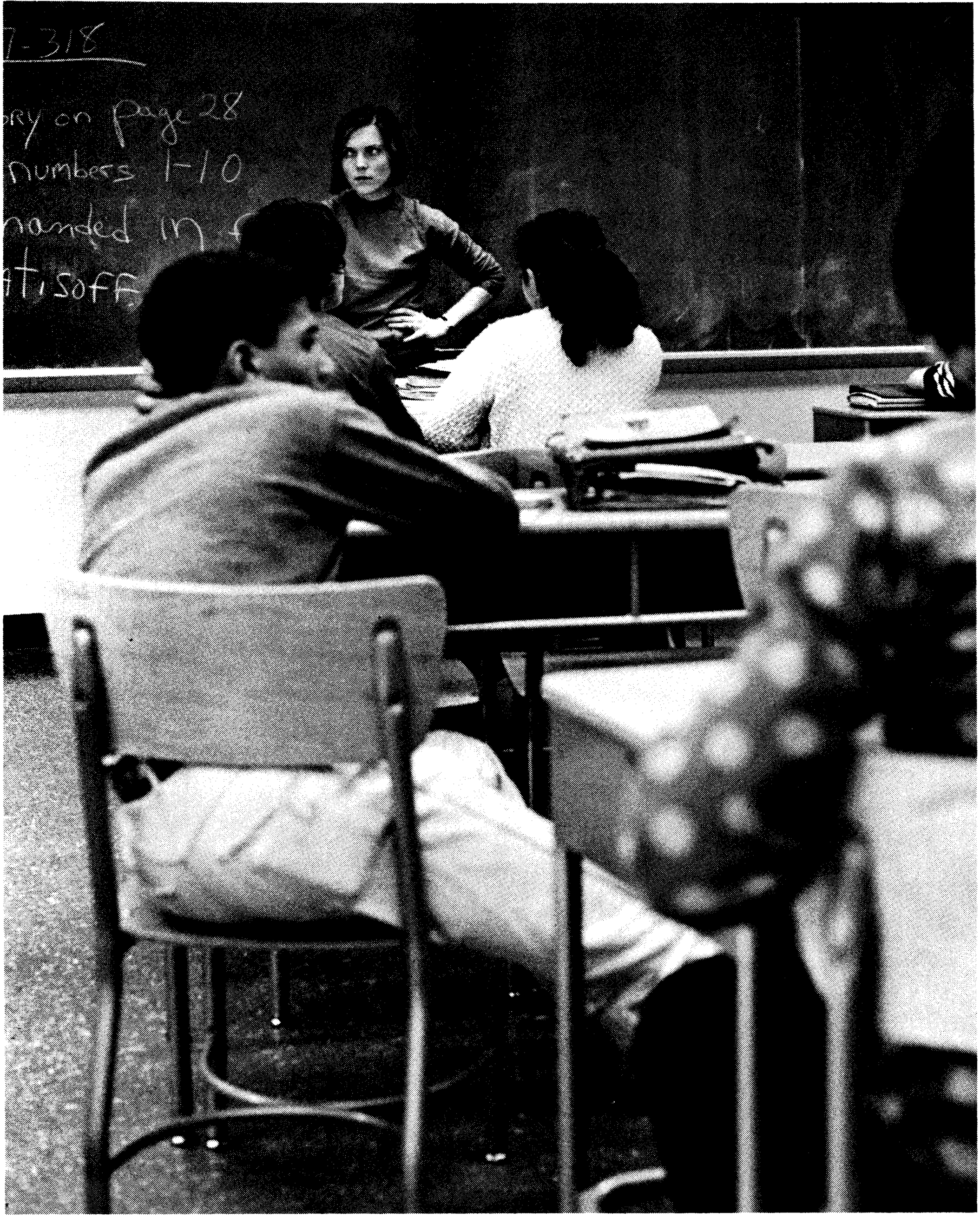
Suppes, Patrick. *Problems of Optimization in Learning a List of Simple Items.* Psychology Series, Technical Report No. 57. Stanford: Institute for Mathematical Studies in the Social Sciences, Stanford University, July 22, 1963a.

Suppes, Patrick. "Mathematical Concept Formation in Children." Unpublished invited address to Division 15 of the American Psychological Association, August 31, 1963b.

Suppes, Patrick, and Ginsberg, Rose. "Application of a Stimulus Sampling Model to Children's Concept Formation With and Without Overt Correction Responses." *Journal of Experimental Psychology,* **63**: 330–36; April 1962.

Suppes, Patrick, and Ginsberg, Rose. "A Fundamental Property of All-or-None Models, Binomial Distribution of Responses Prior to Conditioning, With Application to Concept Formation in Children." *Psychological Review,* **70**: 139–61; March 1963.

Woodworth, Robert S., and Schlosberg, Harold. *Experimental Psychology.* Revised edition. New York: Henry Holt, 1954. 948 pp.



Joe Molnar

PART 5

The Learning Situation and Its Management

5.1 *Human Motivation and Environment*

LLOYD E. HOMME

Reprinted from *The Learning Environment: Relationship to Behavior Modification and Implications for Education. Kansas Studies in Education,* 1966 (June), **16,** 30–39, with permission of the author and *Kansas Studies in Education.* Lloyd E. Homme is with Independent Learning Systems, Inc., in San Rafael, California, and was manager of the research department of the Westinghouse Learning Corporation.

Homme makes a number of points showing differences between human and animal subjects in learning behavior and motivation and then goes on to indicate how operant learning principles can be applied to classroom learning problems in the form of behavior modification. Although behavior-modification techniques generally depend on the learner's being reinforced by an experimenter, or other person who controls the contingencies of reinforcement, Homme recognizes that the end goal is that of developing a learner who is self-reinforcing.

One way to get behavior executed is by the use of punishment or threats of punishment. "Do X or else . . ." appears to be the most popular behavioral contingency in our culture. But control of behavior by aversive stimulation is not the topic of this paper. I will assume that we are in agreement concerning the undesirable side effects of control through aversive stimulation and that what is needed is a system for control through positive reinforcement.

The Conventional Motivation Concept

Conventional motivation concepts appear to be quite satisfactory as long as one stays in the laboratory. The deprivation operations one performs are said to induce a drive or need state which in turn is reflected by a number of physiological and behavioral changes. According to this conception, one change is that the drive state makes a reinforcer available. The reinforcer, of course, is that of which the subject was deprived. For example, when an animal is deprived of water, a thirst is induced which can be reduced by the presentation of water. This is often said to explain why water is a reinforcer, under these circumstances.

The conventional conception of operations and drive can be summarized in the diagram:

deprivation operations —> drive state —> physiological and behavioral changes

Drive is thus a logically impeccable intervening variable firmly anchored in observables at both ends. Drive is said to be a useful construct to summarize the changes which occur as a result of the antecedent operations. This may be true. And, as I mentioned, no particular problem arises from this formulation as long as one remains in the laboratory.

I believe it to be a fact, however, that what is really wanted from the motivation concept is more than is supplied by this intervening variable construct. Evidence for this is found in elementary psychology textbooks. "A motive is whatever moves or incites to action" is a typical introduction to the topic of motivation. The fact that after this introduction the textbook author quickly moves on to the drive construct should not obscure the fact that he really has larger things in mind. One author gives himself away in this fashion: ". . . a drive is an impetus to behavior."

The point is that, as soon as one gets away from the operational laboratory definition of the drive construct, good as it sounds at first, it becomes quite useless. As soon as it is generalized to everyday situations, the necessity for specifying the antecedent operations said to produce it are commonly forgotten. Thus one is left with a topic of conversation rather than a means of manipulating or controlling behavior.

This shows up in its clearest light when motivation constructs are invoked to explain some lack of control of behavior. "Why doesn't Johnny learn?"

Explanation: "He just isn't motivated." This explanation for the lack of control of behavior is rejected on the same grounds that other fictional causes of behavior are rejected. This everyday motivation concept implicitly places the cause for the present behavior inside the organism somewhere. Since it is quite clear that the explainer cannot get at the inside of the organism to fix this, he is freed of the responsibility for control. Through no fault of his, some prior operation hasn't been performed, or the organism's history is somehow deficient.

This objection is only a special case of a general objection to this sort of an explanation. It leads nowhere. It has no consequences in the sense that it does not specify any manipulable variables to control the behavior now.

Further, this tendency to look to the past for variables to control behavior obscures the search for manipulable variables which may exist in the present.

Reversal of a Concept

Usually the motivation problem is conceptualized as follows. If you have a motivated organism, you have a reinforcer. I suggest that we reverse this conceptualization of the relationship between reinforcers and motivation to be: If you have a reinforcer, you can motivate an organism. This formulation has the happy effect of causing one to search for reinforcers rather than bemoaning the lack of motivation. Happily, too, laboratory developments of the last few years tell us where to look.

Premack's P-hypothesis

As we have seen, the usual formulation of motivation involves the drive construct. David Premack, of the University of Missouri, however, has a different formulation which he calls the differential probability hypothesis, or P-hypothesis. He hypothesizes, and is finding considerable laboratory evidence in support, that any higher probability behavior will reinforce any lower probability behavior upon which it is contingent. From this standpoint, there is no need to invoke a drive concept to account for the reinforcing properties of water for a water-deprived animal. The experimenter has arranged matters through deprivation so that drinking is a higher probability behavior than the barpressing behavior which the experimenter wishes to reinforce.

In typical laboratory operant conditioning experiments, matters are arranged so that a single behavior remains at a high probability throughout the experimental session. Thus the same reinforcer can be used again and again. Taking David Premack (1959; 1965) literally, there is no reason to impose this constraint. *Any* behavior can be used as a reinforcer of *any* lower probability behavior *at the instant* it is a higher probability behavior.

Humans and Deprived Laboratory Animals

Nondeprived humans behave differently from deprived laboratory animals. Their behavior differs in many ways, but what is of interest to us now is the wide

fluctuation in behavior probabilities. Rather than having one behavior, such as water-drinking or food-eating, which remains a high probability behavior throughout an extended period of time, just the opposite is the case. With the nondeprived human, specific behavior probabilities vary from very high to very low from one instant to the next. At one moment, getting a cup of coffee may be a high probability behavior. A moment later, taking a sip of it is a high probability behavior; a moment after that, taking a puff of a cigarette is a high probability behavior, and at that instant, taking a sip of coffee may be close to zero probability. Immediately after the puff of a cigarette, the probability of smoking may be near zero and taking a sip of coffee may be quite high again. And so it goes with other responses throughout the day. In general, a response may at one moment be a high probability behavior, then once executed, the probability of immediately repeating it is close to zero. The reason for dwelling on the obvious—that the probabilities of any one behavior typically vary from high to very low from moment to moment—is that, if Premack's P-hypothesis is taken literally, *each* high probability behavior is potentially as good a reinforcer as any other *for the instant* that it remains a high probability behavior.

If the P-hypothesis continues to be confirmed and turns out to be one of the fundamental laws of nature, we have to assume that it is older than man himself. Indeed, once one starts examining the matter, one can see that every one has known about the differential probability principle all along. Grandma knew all about it, for example, when she said, "Clean up your room; then you may go out and play." Grandma, the Contingency Manager, is saying, in effect, "Run off an amount of low probability behavior; then you may execute some higher probability behavior for awhile." However, Grandma was the worst conceivable Contingency Manager when she said, "All right, play one game of cards; then you've *got* to do your homework." Things are exactly reversed here. She is permitting some high probability behavior to occur (playing cards), and then following it with some lower probability behavior (doing homework). A Contingency Manager who knew what he was doing would have these two activities in exactly the opposite order. I hope this one example illustrates the point that it is really quite difficult, at first, to take a scientific behavior principle literally enough to be systematic in applying it.

Contingency Management in the Nursery

As I have previously reported (Homme, et al., 1963), it was the differential behavior probability concept which, when taken literally, enabled us to transform bedlam into a highly controlled nursery school situation. For those of you who have not seen our brief report, I will summarize what happened.

We were faced with the task of controlling the behavior of three three-year-olds without the use of punish-

ment or threats of punishment, implied or otherwise. Furthermore, we had an additional constraint that we not use candy or trinkets as reinforcers. The amount of control exercised on the first day can be summarized: none. One child was running and screaming, another was pushing a chair across the floor (rather a noisy chair), and the other was playing with a jigsaw puzzle. Once our scholars discriminated that punishment did not follow these activities (the rate at which this discrimination was made must have set a new indoor record), the response to the verbal instruction, "Come and sit down now," was to continue the running and screaming, chair-pushing, and so forth.

I have a recurring fantasy about what happens next. I am running alongside the child who is running and screaming and I am screaming too. I am screaming in her ear, "A motive is whatever moves or incites to action!" and "A drive is an impetus to behavior!" Then the child asks, "Is the drive firmly anchored to observables?" I answer, "Yes, certainly." The child continues, "Good I have a running and screaming drive right now which needs reducing. Let's run and scream."

Back to reality.

Almost in desperation, we took Premack's P-hypothesis seriously, and labeled, as high probability behaviors for the children, those behaviors which were aversive to us. We made engaging in these behaviors contingent on the subject's doing a small amount—very small at first—of whatever we wanted them to do. A typical early contingency was merely for them to sit quietly in chairs and look at the blackboard. This was followed almost immediately by the command, "Everybody, run and scream now." This kind of contingency management put us in immediate control of the situation. We were in control to the extent that we were able to teach everything in about one month that we could discover was ordinarily taught in first grade.

Our Contingency Management in the nursery study was quite artistic in that the Contingency Manager (the experimenter who decided what behavior was going to be contingent upon what behavior) made his decisions on the spot. Following this, we were forced to get less artistic, more specific.

The First HEW Contract Group

Under contract to the Office of Education, Department of Health, Education, and Welfare, we were given the chance to teach literacy to 26 school dropouts, or potential dropouts. Since each student (*S*) had different high probability behaviors, we were forced to specify in advance what the contingencies were going to be for each *S*. This evolved into what we now call Contingency Contracting.

Each day when *S* arrived, he found a Contingency Contract awaiting him. The Contingency Contract specified a series of pairs of behavior, each pair consisting of an amount of lower probability (task) behavior leading to a set time of a higher probability (reinforcing) behavior. Besides the left-to-right or within-pairs progression of probabilities, there was a general progres-

sion between pairs from lower to higher probabilities. The *S* started each day's work, ideally, with the lowest probability behavior and ended it with his highest. For example, if the *S* "hated spelling," the first pair of activities for the day might have been: three pages of spelling leading to five minutes of the higher probability behavior (HPB) of a coffee break. Toward the end of this contract, the task of a pair of activities would be something he "didn't mind doing too much." For example, it might have been five pages of reading followed by the higher probability behavior of ten minutes' discussion of what was read.

We were quite impressed with our motivation results. To appreciate our enthusiasm, one must realize that these *S*'s had, like other dropouts, or potential dropouts, histories of punishment for absenteeism and other nonacceptable behavior, for failure, and for other aversive events associated with school. What is more, they looked and acted the tough teenager role. Some of them wore dark glasses throughout the six-weeks' experiment even though they worked indoors. We discovered some were carrying knives.

During the first week, while we were still gearing up to utilize Contingency Contracting, we had three dropouts out of the 26 students enrolled; following that first week, there were zero dropouts. Attendance was virtually perfect. Even more impressive, but difficult to document, was the emotional change observed. Smiles became more frequent. Tasks (short ratios which led to a high probability behavior) were attacked eagerly. Just as impressive and easily documented was that, despite the fact that *S*'s came from neighborhoods in which a special kind of diplomacy—fighting—is used to settle disagreements, there were no fights on our project. So far as we know, there were not even any arguments among the students. A remarkably serene atmosphere prevailed.

The academic gain achieved over the six weeks' period was disappointing to us. We effected, according to the California Achievement Test, a mean gain of .5 grade level in six weeks' time. Part of the reason for this showing was a lack of control of the quality of the lower probability behaviors. To correct this, we now have more Progress Checks, miniature tests which are part of the lower probability behaviors. These are now routinely included in task assignments. For example, a spelling task assignment does not read "Five pages of spelling," but "Five pages of spelling plus Progress Check."

Another reason for getting only an average of .5 grade level gain in six weeks was that several *S*'s scored lower on the posttest than on the pretest. Incredible as it may seem, they apparently lost ground while working so hard. When it was all over, we learned how it happened that some *S*s' scores went down, rather than up. They had cheated on the pretest and not on the posttest.

The Second HEW Contract Group

In our enthusiasm for the success of Contingency Contracting, we assumed that teachers would be eager to adopt the system, and that it would be simple

to teach them how to use it. As we look back, we can see that we were incredibly naive, that we grossly underestimated not only the inertia but the active resistance to change on the part of at least some teachers. Our plan was to have the teachers read eight or ten pages of material (Homme, 1964). Nobody read them. Then this was to be followed by three half-hour sessions of lecture and discussion of the system. Nobody listened. They had "had all this stuff in college psychology class." Besides, they had been working with juvenile delinquents for some time. (How long had *we* been working with juvenile delinquents?) They knew that what was required was punishment, not this positive reinforcement stuff.

To sum up the results with the second group of students, the system was so compromised and misused that no posttesting was done on the grounds that the results would have meant nothing.

That is not to say that nothing was gained from Group 2. On the contrary, we learned that our plans for communicating with the teacher were entirely inadequate. We had planned to write a small manual based on the three lecture-discussion sessions we had with the teachers and on the material we had previously written. This was going to constitute the teacher's manual. This experience with Group 2 quickly and thoroughly convinced us that our plan simply would not work. In its stead, we prepared a detailed, step-by-step utilization manual (Homme, 1965) written in the style of "This is the way it's done. Do not deviate from this plan."

The Third HEW Contract Group

A manual such as the one described has been prepared and is currently in use in a Special Education class. Contingency Contracting is being used, and things appear to be going quite well. The teacher is quite enthusiastic about what is happening, but there is an assistant principal who is giving off some portentous signals. She is looking into the project. What this will mean for the project's future, and whether some administrative reason will be discovered to disrupt the present smoothly-running operation, only time will tell.

The Learning Center

In addition to the rest of our activities, we are also taking individual referrals from local guidance counselors, psychologists, and the like. We have students ranging from borderline retardates to those of superior intelligence. At one end of the achievement scale are those students having trouble in school; at the other end are the bright students who want to get even farther ahead. Mostly, however, our clients are students who are having trouble in school.

Across the board, with the exception of the superior students, our main job is to motivate the client. We motivate students in a somewhat peculiar way. In analyzing the circumstances under which a person will or will not say another human is motivated, we have concluded that his reaction will depend largely upon the display of three factors by the *S*: (1) the speed with which

he works (or the latency of task attack); (2) the frequency with which he smiles; (3 certain other verbal behaviors (e.g., "I'm learning a lot." "I like it here.")

With Contingency Management, we deliberately condition these symptoms of high motivation. To get speed, we superimpose a differential reinforcement of high rate (DRH) schedule on the fixed ratio task schedule. This simply means that we say, "Do 25 frames of spelling. If you do them in less than 20 minutes and pass your Progress Check, some good contingency will befall you." After this kind of conditioning, the subjects appear to be highly motivated humans. We have no idea whether they "really" are or not. Furthermore, if they continue to display these symptoms, we don't care.

I have always been impressed with the operant conditioner's control over the attention span of an ordinary pigeon. I don't know what would happen, exactly, if educators were asked to measure a pigeon's attention span, but I suppose they would conclude it was quite short. I also suspect that the operant conditioner wouldn't pay any attention to it. He would hold the pigeon's attention as long as necessary—up to 15 hours or so if he wanted to.

To see whether we could come anywhere near approximating this with humans, we have, on two occasions, kept an *S* responding for a full working day, from 9:00 A.M. till 5:00 P.M. with time out for lunch. (The latter was, of course, contingent on executing some lower probability behavior.) Both of these sessions produced gains of better than a year's grade equivalent. We are certain that we can routinely do *at least* this well. We have started referring to this, incidentally, as the "one-day fix." We don't do this oftener because of the manpower it requires—one human contingency manager per *S*.

Westinghouse's SLATE, the Computerized Contingency Manager

The manpower requirement I mentioned will soon get remedied. It will get remedied by SLATE[2], the first real teaching machine I have ever seen. Each student (*S*) will be seated at a console containing a keyboard, a microphone, earphones, and a cathode ray tube for output display. The cathode ray tube also serves as a stimulus area to which the *S* can respond by means of the keyboard or by means of an electronic pencil with which he can be asked to point at various objects displayed on the tube or various areas of the display. I could go on talking about SLATE all day without exhausting the topic of what this almost magical machine can do, but I will summarize: everything. Specify the function you want, and I can guarantee that SLATE can do it. That is not quite true. It cannot dice carrots, and it does not have an auditory discrimination capability—that is, it cannot read human speech. But, allowing for these restrictions, I'm convinced can do everything.

Here is how the classroom of the future will work. The student will sit down at his console and will announce

[2] This equipment is called SLATE, an acronym for Sensitive to Learner Automated Teaching Equipment.

his arrival by typing his name. (If he is too unskilled to do that, we can have a monitor do it, or arrange for some easier signal to be given.) After that, SLATE will take over. It will remember what contingencies work best, with what latencies *S* responds, and what concepts he had trouble with yesterday (or last month, if you like). When you consider that a single computer can handle a hundred student consoles simultaneously, each presenting a different program, you get some idea of the power of this educational tool.

SLATE Simulators

While awaiting the completion of a final SLATE model, we have been preparing for it by writing programs especially for SLATE using humans to simulate computer operations. Let me tell you that humans generate only a poor approximation.

A simulator in action might proceed something like the following. He might say, "Puh," and *S* would respond simply by using a pencil (rather than an electronic pointer) to point to a *P*. The simulator then would arrange a new display, utter another sound, and so on. This sort of simulation is not too bad, but it is far from perfect. It is far too slow for us to learn anything about maximum rate, for one thing. Recently we have been working with two preschool slum kids, supposedly difficult organisms to motivate. We find this is simply not so. These *Ss* learn like all other organisms. They are being run only about three hours a day at present, and the reason for not running them for a greater period of time is interesting. It is not that the *Ss* get fatigued; it is the SLATE simulators who wear out.

The High Probability Behavior Area (HPB Area)

As I have indicated, many high probability behaviors involve considerable action and generate a good deal of noise. This being the case, we have found it necessary to set aside a special area for HPB's to take place. As you may suspect, a supervisor is needed in this area; we call the supervisor the HPB Administrator.

The way the system works is this. The *S*, interacting with the SLATE simulator, runs off the ratio specified by the contract, e.g., pointing to the letters appropriate for a half-dozen sounds. At the end of this ratio, *S*, is handed a card which specifies an amount of contingent time, say two minutes. *S* grasps this card and runs to the HPB area, hands his card to the HPB Administrator and sets his timer. He then begins his coloring, dart throwing, talking with the HPB Administrator, or whatever else the HPB Administrator specifies.

When the bell on the timer rings, signifying that contingent time has elapsed, the subject races back to the task area, and begins working again. If it seems surprising that the subject races to return to the task area, let me assure you that it was not always thus. When we first started on this project, the *Ss* would run to get to the high probability area all right, but they would dawdle for what seemed an interminable time in returning to the task area. Laboratory animals on a fixed ratio schedule often show a similar behav-

ioral phenomenon. After a reinforcement, they pause for varying lengths of time, depending on the richness of the schedule. Undue pausing is common enough so that it is said to reflect "ratio strain." This pause after reinforcement can be eliminated by shifting to a variable ratio schedule. This is what we did with our subjects, in effect. After returning to the task area, they would intermittently be presented with the instruction to return immediately to the high probability area. A further contingency was superimposed on this schedule: The instruction to return to the high probability area was given only if the return to the task area was prompt. With these contingencies, our subjects are now running so fast, rounding corners with such lightning speed, that we are in some fear that they might hurt themselves by their excited slipping and sliding and bumping into SLATE simulator's chairs.

The Future

No one can say with any certainty what the future will bring, of course. Nevertheless, guesses about the future are entertaining and sometimes more.

I think it is quite obvious from what has already occurred that an area which is going to yield exciting and important results is that of self-management of contingencies. A corollary of this is that it will be necessary to develop a technology for teaching the self-mangement of contingencies. It is not known how young this teaching may begin, but there are some encouraging signs. At least one five-year-old in the world, whose mother has been playing the Contingency Management game with her, is now beginning to play the game by herself. One day the five-year-old was observed opening the door of the freezer and beginning to reach for a popsicle when she suddenly paused. "Whoops!" she said. "I have to make my bed first." At this she dashed off, made her bed, and dashed back to tthe freezer. "Now I may have my popsicle," she said.

No one has yet seen an adult human who started life early as a self-contingency manager. But the chances are the results will be better than anticipated.

There is very little doubt that self-management of contingencies can be taught. Questions which remain are how early it can be taught, how much of it can be taught, and how fast it can be learned.[3]

One thing remains certain. No one knows how fast a human being can learn *anything*; we have not yet begun to approach the limits of the human's capacity to learn. And we will never begin to approach them until a computerized contingency manager such as Westinghouse's SLATE is properly programed and used.

References

Homme, L. E. Perspectives in psychology: XXIV Control of coverants, the operants of the mind. *Psychol. Rec.,* 1965, **15**, 201–511.

Homme, L. E. Technical progress report no. 1:

[3] In a paper published since this symposium was held, the principles of self-management have been extended to control the frequency of events ordinarily called mental (coverants). Preliminary indications are that this kind of control can be learned without difficulty by young children.

a demonstration of the use of self-instructional and other teaching techniques for remedial instruction of low-achieving adolescents in reading and mathematics, submitted to U.S. Office of Education under Contract No. OE-4-16-033 October 1, 1964.

Homme, L. E. Technical progress report no. 2: a demonstration of the use of self-instructional and other teaching techniques for remedial instruction of low-achieving adolescents in reading and mathematics, submitted to U.S. Office of Education under Contract No. OE-4-16-033 January 1, 1965.

Homme, L. S., C. deBaca, P., Devine, J. V., Steinhorst, R., and Rickert, E. J. Use of the Premack principle in controlling the behavior of nursery school children. *J. Exp. Anal. Behav.,* 1963, **6**, 544.

Premack, D. Reinforcement theory. In Levine, D. (Ed.)., *Nebraska Symposium on Motivation 1965.* Lincoln: U. of Nebraska Press, 1965.

Premack, D. Toward empirical behavior laws: I. positive reinforcement. *Psychol. Rev.,* 1959, **66**, 219–233.

5.2 *Negative-Attention Seeking or Tactical Provocation?*

RONALD GALLIMORE

This unpublished paper (1969) is reproduced here with the permission of the author. Ronald Gallimore is a member of the psychology and anthropology departments of the University of Hawaii and has done considerable research in social learning, educational processes, and cross-cultural factors in personality.

The paper deals with a perennially interesting subject: the attention-getting value of misbehavior. Gallimore prefers to refer to such acts as "tactical provocations," which are, as he says, learned ways of responding to problem situations. Students display such provocations in such a way that teachers are tricked into reinforcing them. As a result, disruptive or other types of negative behavior assume a high priority in the learner's repertory of responses, and more positive forms fade into the background.

It is often suggested that some children deliberately provoke negative adult attention in the absence of positive reactions. Such children are said to be negative-attention seeking, presumably because their behavioral repertoires are for one reason or another inadequate to the task of eliciting positive reactions. Some observers have gone so far as to suggest that such individuals exhibit a "need" to be punished (Josselyn, 1958), while others have argued that children simply "need" attention, even if it is negative (Davis, 1949). Social learning theorists, on the other hand, assume that what an adult may regard as negative-attention seeking is, in many instances, the consequence of an intermittent reward and punishment schedule which is not apparent to the observer (Bandura and Walters, 1963). That is, the observer incorrectly infers that a child *intends* to provoke negative attention. To illustrate, a preschooler may persist in disrupting the teacher-led activities of his classmates despite the fact that he is often reprimanded for such behavior—what maintains the behavior is the reinforcing effect which occurs, perhaps while the teacher is absent or distracted, when his antics provoke a positive peer-response. Thus it is unnecessary to assume either that the child needs or enjoys the negative reaction of his teacher. Indeed, the use of the term in such cases is descriptively ambiguous, and theoretically unnecessary.

Bandura and Walters (1963) also suggest that many responses labeled negative-attention seeking may be examples of anxiety-motivated responses which are often highly resistant to extinction. An individual may persistently expose himself to aversive, or negative attention, in his efforts to escape more unpleasant consequences. Such a conception has been, of course, widely accepted as an explanation for the persistence of self-defeating, neurotic behavior.

Whatever the attitude of psychologists, negative-attention seeking has continued to be a popular explanatory concept as well as a descriptive category. For example, in a recent novel written apparently by an experienced teacher (*Up the Down Staircase,* Kaufman, 1964), there are a number of references to students whose unruly and delinquent behaviors are classified as substitute means of gaining attention which they have failed to secure with more socially approved strategies. In one subplot, it is made explicit that a bright, but alienated boy, deliberately provokes negative reactions as a preliminary tactical maneuver, designed to establish a positive relationship with the heroine. Of course, his behavior could also be regarded as defensive, or aggressive; or, for that matter, clever since he does succeed in eliciting highly positive attention from the teacher-heroine in a school staffed otherwise by hostile or indifferent personnel. Once the positive relationship is established—that is, the teacher-heroine begins to dispense positive reinforcers, e.g., approval of the boy's articulate protest against an authoritarian school—there is a sharp reduction in the frequency of negative-attention seeking behaviors. At least the central character perceives a reduction in relation to herself, which may simply reflect a

recategorization effected by a better understanding of the boy's intentions. From a social learning view, of course, there is still no requirement for a negative-attention seeking explanation since the boy is reinforced for a negative-positive response sequence and probably on an intermittent schedule as new teachers enter the school. It is quite reasonable to assume, however, that some children are trained, unintentionally no doubt, to initiate interactions by provocative acts. Such tactics may serve as a means of attracting attention to subsequent responses to which adults typically respond with positive reinforcers. While such sequences can be explained without reference to a negative-attention seeking concept, as a descriptive category it may retain some usefulness.

Such a position is supported by a recent laboratory study (Gallimore, Tharp, and Kemp, 1969) which has suggested that for children with strong attention-seeking habits, exposure to a period of social isolation will increase their preference for mild adult reproof relative to no attention at all. Similarly, MacDonald and Gallimore (in press) offer a number of examples of apparently "negative" teacher attention serving to maintain undesirable student behavior. Such findings have also been reported by Tharp and Wetzel (1969).

Aside from the problem of distinguishing theoretically between punishment and negative attention, there are at least two complex conceptual issues relevant to an adequate definition. First is the problem of intent, an issue which plagues most attempts to define response categories—for example, if a child strikes his mother's knee, is he (1) being aggressive, or (2) merely making a strong dependency overture, which presumably could also be interpreted as an example of negative-attention seeking?

The second issue is the question of individual perception. That is, can adults assume their interpretation of the meaning of social acts to be identical to the interpretations imposed by children? And vice versa.

Negative-Attention Seeking or Tactical Provocation?

In spite of the theoretical problems which negative-attention seeking creates as a descriptive category and as an explanatory concept, it will undoubtedly continue to enjoy broad usage in applied and lay settings; and this support may be deserved in view of the usefulness of the idea that some children are able to secure positive attention only after they have elicited negative attention. A provocative act with some liklihood of eliciting adult reaction, perhaps initially unpleasant but subsequently appealing, is probably a tolerable risk for children with few viable alternatives for securing positive attention. For those with strongly established attention-seeking response patterns, the liklihood of such sequences is probably greater.

Perhaps the term *tactical provocation* conveys more accurately and fully the potential behavioral referents of what has been called negative-attention seeking. Tactical provocation can encompass both those instances in which the long-range goal (strategy) is elici-

tation of positive reactions and those in which the individual, for some strategic reason, seeks to provoke a negative reaction. In the first category, the child whose parents have inadvertently encouraged provocative responses, by ignoring positive overtures for attention, serves as an illustrative example: indeed, Sears, Maccoby, and Levine (1957) conclude that extreme dependency, which parents may describe as irritating, is generally behavior which was freely and amply rewarded at some earlier age. The child, according to Sears *et al.* (1957), responds to the gradual withdrawal of reinforcement with intensified effort—that is, he does more of what was once rewarded, and if his parents occasionally respond to these overtures, the pattern will persist and the child may be regarded as a negative-attention seeker. Or, if you will, he becomes a provocateur. It is likely that such training sequences as these are involved in a majority of those instances in which the child is described as a negative-attention seeker.

The provocation of negative attention may also be a strategically useful tactic for adults. For example, a political candidate may gain increased public sympathy if he can provoke his opponent into making an intemperate, personal attack. In such cases, of course, the individual is seeking attention outside the dyadic relationship in which the tactical provocation occurs. In some cases the goals of the provocateur may be within the dyad; for example, if an individual attaches personal significance to engaging in and winning emotionally-charged arguments, he may pursue this goal by raising for discussion topical areas in which he has special competence or access to information not available to others. After provoking the target person into making a statement(s) which the provocateur knows to be false, because of his special resources, the *coup de grâce* can be delivered; whatever value such exchanges possess is undoubtedly an individual matter. It is also likely, however, that the intensity or seriousness with which provocative tactics are employed are important variables, and are a function of many factors. If other people are present, or involved, as in the example of the political candidate, the stakes of the game are likely higher.

There are several advantages to use of the term tactical provocation as a descriptive category in which to include those behaviors generally called negative-attention seeking. First, it is more clearly a simple, descriptive label and thus may be less likely to be employed as an explanatory concept—that is, although the term negative-attention seeking is, as suggested, a potentially useful descriptive category, it is frequently and inappropriately used to "explain" behavior in a completely circular manner, just as such terms as schizophrenia are misused.

Second, the proposed change is consistent with a recent and persuasive proposal for the use of the term *tactic* as an alternative to the concept of defense mechanism (Sechrest and Wallace, 1967). Sechrest and Wallace define tactics as "plans for dealing with relatively immediate and limited problems, and they usually form a part of a larger strategy. . . . It is often difficult to distinguish between a tactic and a

strategy, but we will be satisfied here with the idea that tactics are plans that are *limited* both in time and scope while strategies involve longer term plans that are more pervasive" (Sechrest and Wallace, 1967, pp. 305–307).

Although they make a strong argument for the concept of tactic, as opposed to defense mechanisms, they add: "It should be kept in mind that we are not making any assumption that tactics as such really exist. . . . We merely want to attempt to view behavior from that standpoint to see whether it is meaningful and helpful to us. But, so far as we know, it is we, not the ordinary man, who think in terms of tactics. It is more or less convenient fiction like so many other psychological constructs" (Sechrest and Wallace, 1967, p. 306).

It is further assumed that tactics are learned responses, acquired in ways accountable by social learning principles and that, in general, are patterns of response elicited in problem situations. The authors add, however, that they "do not wish to rule out the possibility that various tactics might be 'invented' without prior trials to meet particular problems" (Sechrest and Wallace, 1967, p. 307). The greater congruence of tactical provocation than negative-attention seeking, with learning theory can be probably considered a third and definite advantage.

A fourth and particularly appealing quality of the term itself is its relative immunity to psychopathological connotations. As Sechrest and Wallace suggest, most tactics "may well have their place, their appropriateness, depending upon the capabilities and limitations of the individual, and on the situation in which he finds himself" (Sechrest and Wallace, 1967, p. 366). In other words, few responses are absolute indices of maladjustment. Thus, labeling the behavior of a child, or an adult, as tactically provocative may or may not imply pathology, immaturity, or, indeed, anything negative.

Fifth, the suggested use of tactical provocation makes more apparent the relevance of a wide variety of theory and research to the behavioral referents of the term. For example, the integration studies of Jones (1964) are quite clearly related to the phenomenon, as well as the burgeoning studies of power and influence in dyadic interactions (Jones and Gerard, 1967). An example of the value of introducing the term, tactical passivity, is particularly well illustrated by Walters and Parke (1964). Walters and Parke were able to bridge the gap between various research areas, developmental, social, personality, learning, by proposing that the apparent unitary nature of dependency, which had been a traditional variable of interest in developmental investigations, be considered a function of attending and orienting behavior. By this reconceptualization, they were able to offer an integration of traditional investigations of dependency, research on the effects of nurturance and social isolation, and affiliative behavior among other major areas of psychological research.

And, finally, the problem of inferring intent. Is tactical provocation less susceptible to that problem than negative-attention seeking? The adage, one man's meat is another man's poison, makes the point more succinctly but no less sharply than the observation that

individual, social, and cultural variations in the definition of reward and punishment, success and failure, pleasure and pain, often involve dramatic alternations in the classification of particular stimulus events (Guthrie and Jacobs, 1966; Sechrest and Wallace, 1967; Walters and Parke, 1964). In the final analysis, behaviors labeled negative-attention seeking, or tactical provocation, must rest on the ethno- and egocentric inference that what is poison to the observer is poison to others. Such problems question how clever our science can become.

Tactical provocation appears to be a popular pastime among academicians, especially prevalent at national and regional meetings. Perhaps provocative tactics are healthy and necessary within the academic community although at least one of our illustrious forebears suggests otherwise, at least when the exchange is public. Sir Isaac Newton wrote:

"What's done before many witnesses is seldome without some further concern then that for the truth: but what passes between friends in private usually deserves ye name of consultation rather than contest, & so I hope it will prove between you & me" (Merton, 1965, p. 23).

In commenting on Newton's remarks, Merton (1965) quotes a relevant passage from some of his earlier work:

"I suggest that often these polemics have more to do with the allocation of intellectual resources than with a closely formulated opposition of sociological ideas."

"These controversies follow the classically identified course of social conflict. Attack is followed by counter-attack, with progressive alienation of each party to the conflict. Since the conflict is public, it becomes a battle for status more nearly than a search for truth" (Merton, 1959, pp. 29–30).

Implications for Teachers

The theoretical delights which the controversy may provide psychologists are of little interest to teachers, unless from the discussions comes something instructive. The value of clarifying the theoretical issues involved in the preceding comments can be best demonstrated by reporting on some recent advances in the management of children who exhibit tactically provocative behavior. In general the advice is as follows: first, assess the relative frequency of reinforcement for provocative behavior as against acceptable behavior and do not be surprised to find that, in most instances, the teacher is unwittingly encouraging that which she despises (Tharp and Wetzel, 1969). If such is the case, advise the teacher to cease immediately reinforcing behavior she does not wish to occur and begin to reinforce what she regards as acceptable. As simple as the remedy appears, it is by no means obvious to the harried teacher who must divide her attention among many children—she simply has little time to discover the contingencies between the behavior of tactically provocative children and her own reactions. In our experience (MacDonald and Gallimore, in press) in Hawaii, we have often found teachers stunned by the demonstration that the amount of "bad" behavior exhibited by

their charges is directly proportional to the amount of "cease and desist" behavior on the part of the teacher herself. Thus, while it may entertain some to attribute provocative behavior to a mysterious "internal" mental entity or to deplorable home conditions, it is more in keeping with the evidence to assume that "bad" behavior is a function of the environment in which the "bad" behavior occurs. In rather ironic fashion, it does seem true that "good" behavior needs attention but this insight should not obscure the basic fact that it is the responsibility of the teacher to manage the child's environment in such a way as to insure that attention is contingent on good rather than provocative behavior.

References

Bandura, A., and Walters, R. H. *Social Learning and Personality Development.* New York: Holt, Rinehart, and Winston, 1963.

Crandall, V. C. Personality characteristics and social achievement behaviors associated with children's social desirability response tendencies. *Journal of Personality and Social Psychology,* 1966, **4**, 477–486.

Crandall, V. C., Crandall, V. J., and Katkovsky, W. A children's social desirability questionnaire. *Journal of Consulting Psychology,* 1965, **29**, 27–36.

Crowne, D. P., and Marlowe, D. *The Approval Motive.* New York: John Wiley, 1964.

Davis, K. *Human Society.* New York: Macmillan, 1949.

Dorwort, W., Ezerman, R., Lewis, M., and Rosenhan, D. The effect of brief social deprivation on social and nonsocial reinforcement. *Journal of Personality and Social Psychology,* 1965, **2**, 111–115.

Gallimore, R., Tharp, R., and Kemp, B. Positive reinforcing function of negative-attention. *Journal of Experimental Child Psychology,* 1969, **8**, 140–146.

Gewirtz, J. L., and Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology,* 1958, **56**, 165–172.

Guthrie, G. M., and Jacobs, P. J. *Child Rearing and Personality Development in the Philippines.* University Park, Penn.: The Pennsylvania State University Press, 1966.

Jones, E. E. *Ingratiation.* New York: Appleton-Century-Crofts, 1964.

Jones, E. E., and Gerard, H. B. *Foundations of Social Psychology. New York: John Wiley,* 1967.

Josselyn, I. M. *Psychosocial Development of Children.* New York: Family Service Assoc. of America, 1948.

Kaufman, B. *Up the Down Staircase.* Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Lewis, M., Wall, A. M., and Aronfreed, J. Development change in the relative values of social and nonsocial reinforcement. *Journal of Experimental Psychology. 1963,* **66**, 133–137.

MacDonald, S., and Gallimore, R. *Battle in the Classroom.* New York: International Textbooks, in press.

Merton, R. K. *On the Shoulders of Giants,* New York: The Free Press, 1965.

Sears, R. R., Maccoby, E. E., and Levin, H. *Patterns of Child Rearing.* Evanston, Illinois: Row, Peterson, 1957.

Sears, R. R., Rau, L., and Alpert, R. *Identification and Child Rearing.* Stanford, California: Stanford University Press, 1965.

Sechrest, L., and Wallace, J. *Psychology and Human Problems.* Columbus: Charles E. Merrill, 1967.

Tharp, R. G., and Wetzel, R. *Behavior Modification in the Natural Environment.* New York: Academic Press, 1969.

Walters, R. H., and Parke, R. D. Social motivation, dependency, and susceptibility to social influence. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology, I.* New York: Academic Press, 1964, 231–276.

5.3 *Social Reinforcement in Reducing Inappropriate Behavior*

ALAN S. BRISKIN AND WILLIAM I. GARDNER

Reprinted from *Young Children,* 1968, **24**, 84–89, with permission of the authors and *Young Children.* Copyright © 1968, National Association for the Education of Young Children, 1834 Connecticut Ave., N.W., Washington, D.C. 20009. Alan S. Briskin has been an elementary teacher in the Milwaukee public schools and is now a faculty member of the Department of Counseling and Student Personnel Psychology, of the University of Minnesota. William I. Gardner is a child clinical psychologist and is the chairman of the Department of Studies in Behavioral Disabilities.

In the preceding article, Gallimore discussed some of the ways in which children's misbehavior can be reinforced by adults, and in the first article in this section Homme described ways in which behavior modification can be used in instruction. If cognitive skills can be taught through behavior modification, it follows that forms of social behavior can be taught as well. In the present article, Briskin and Gardner describe how behavior modification procedures were used to socialize the hyperactive, disruptive behavior of a three-year-old girl in a nursery-school setting.

Various behavior change procedures are available to the nursery school educator in dealing with inappropriate disruptive behavior of children in a group educational setting. Of those based on learning principles, time out from positive reinforcement and simultaneous positive reinforcement of alternative behavior which is incompatible with the disruptive behavior provide an approach for the teacher which can be implemented without excessive expenditure of time or personnel involvement.

Taking the child out of the classroom, or "time out" as a form of punishment, does not suppress behavior permanently but serves to increase the chance of obtaining appropriate behavior that is incompatible with the maladaptive behavior. When it occurs, the alternative behavior is reinforced so that its probability of emission will become greater than that of the maladaptive behavior. The crucial element is that an alternative response is available in the subject's behavioral repertoire for obtaining reinforcement.

The use of time out with preschool children has been limited to laboratory situations (Baer, 1960, 1961, 1962a, 1962b) or applied to severe behavior cases including autistic children (Hewett, 1965; Wolf, Risley & Mees, 1964; Risley & Wolf, 1967; Wolf, Risley, Johnston, Harris & Allen, 1967). Furthermore there is a paucity of studies describing successful modification of a child's multiple set of problem behaviors in the field situation. One of the few available, that of Patterson & Brodsky (1966), describes a modification program successfully used to alter a set of inappropriate behaviors in a five-year-old child. A time out procedure was used to alter the subject's assaultive behavior and positive reinforcement to increase cooperative behaviors.

The use of positive reinforcement to shape appropriate behavior in nursery school children is well established (Allen, Hart, Buell, Harris & Wolf, 1964; Harris, Johnston, Kelley & Wolf, 1964; Hart, Allen, Buell, Harris & Wolf, 1964; Johnston, Kelley, Harris & Wolf, 1966). Results from these studies indicate that positive reinforcement can be effectively used in field conditions to shape or change the various behaviors under consideration.

The purpose of the present study was to evaluate the usefulness of a time out procedure along with nonattention in reducing the frequency of occurrence of a complex set of inappropriate behaviors of a nursery school child and social reinforcement to increase the rate of response of socially acceptable behaviors. The study was completed in a field situation where no environmental modification was attempted other than the specific strategy applied to the subject under study.

Method

Subject

Lisa, three years and three months at the beginning of the study, was one of 18 children enrolled in a cooperative nursery school. The class was staffed by an experienced teacher and the mothers of the 18 children who rotated their services over the year.

Lisa lived with her parents and a 14-month-old sister. Reports from parents

and pediatrician indicated normal birth history and physical development. These reports did indicate a history of high activity level and management problems. Observation of the child in the school situation revealed an extremely hyperactive, disruptive, and difficult to control child. Her inappropriate behavior at the beginning of the study was seriously interfering with the normal classroom routine.

Procedures

On the basis of classroom observations and conferences with the teacher and the mother, socially acceptable behaviors in the classroom were operationally defined as: attending to a play activity or assigned work task for more than five seconds, sharing play space and toys with other children, verbally and physically interacting with the other children in a nondisrupting manner, responding to the teacher's verbal instructions, and evidencing self-help skill, e.g., putting away toys, her rest mat, and putting on her coat. Inappropriate behaviors were defined as: screaming and throwing things in fits of anger, crying or whining when not getting her own way, not waiting her turn to engage in art projects and physical activities, biting, hitting, grabbing, rough pushing, leaving the room, group, or activity without reason or permission, and not responding to verbal instructions.

The daily schedule of the nursery school, in session three mornings each week, was divided into seven activity periods. Free play, arts and crafts, snack time, and outdoor play were judged to be relatively unstructured activity periods during which Lisa's inappropriate behaviors were less disruptive of class routine. Story, rest, and music time were structured activity periods during which Lisa's inappropriate behaviors often completely disrupted classroom routine. Inasmuch as a significantly greater amount of disruptive behavior was observed during structured activity periods than unstructured periods (See Table 1) a differential treatment strategy was followed to reduce the rate of occurrence of inappropriate behaviors. A time out procedure was used during the three structured activity periods and nonattention during the four unstructured periods. Socially acceptable behaviors were strengthened by making the teacher's attention and praise contingent upon the child acting appropriately during the seven activity periods.

The child's mother administered the time out procedure, removing Lisa from the classroom upon a signal from the teacher or the experimenter that one or more of the defined inappropriate behaviors had been observed. The utilization of a preschool child's mother in the behavioral engineering process had been established by Wahler, Winkel, Peterson & Morrison (1965) and Allen & Harris (1966). The child was seated outside the classroom for two minutes. The only verbal interaction during the time out was a short statement by the mother telling the child why she was taken out of the classroom. During the four unstructured periods of the morning, the teacher and the helping mother ignored Lisa's inappropriate behavior. When serious altercations occurred be-

TABLE 1. Average Percent of Time Engaged in Disruptive Behavior

Activity Periods	Modification Stage			
	Stage I Baseline (Percent)	Stage II 1st Treatment Period (Percent)	Stage III Reversal (Percent)	Stage IV 2nd Treatment Period (Percent)
Four Unstructured Activity Periods (Nonattention and Teacher Praise)	14	6	4	2
Three Structured Activity Periods (Time Out and Teacher Praise)	54*	5 (a)	4	2 (b)

*$x^1 = 63.51$, $p < .001$ indicating a significantly higher percent of time engaged in disruptive behavior during the structured activity periods. (This statistical association was not indicated in Stage II, III, or IV—$p > .05$.)
(a) Number of time outs: 24. (b) Number of time outs: 11.

tween Lisa and the other children, the teacher and her aides intervened but attended only to the other children.

The teacher had the responsibility of reinforcing the child and carrying out the procedure in a consistent, systematic manner. Although a definite schedule of positive reinforcement was not followed, teacher praise and attention was observed to be consistent and regular in presentation.

The study was carried out in four stages. In Stage I, (Baseline) Lisa's behavior was observed for six school sessions before modification procedures were begun to ascertain the operant level of appropriate and inappropriate behavior in the seven activity periods.

Stage II consisted of the first treatment period lasting for six school sessions. During this stage the modification strategy was differentiated for application during the structured and unstructured activity periods.

During Stage III, treatment was discontinued and inappropriate behavior was treated as it had been prior to the initiation of the study. This stage covered three school sessions.

In the final stage contingencies established during the first treatment period were re-established for six school sessions. During this stage the teacher was observed to be positively reinforcing the subject consistent with the amount of attention given to the other children in the class.

Recording the data

Two observers using a three-second glance/12-second record method gathered data simultaneously and continuously over the 21 days of the study. Seven five-minute time samples were taken each school session (9 A.M.–11:30 A.M.) corresponding to the seven activity periods regularly scheduled.

Each observer recorded 140 observations per school session resulting in 2,940 for the study. Observations were recorded on a behavior data form under the categories Appropriate or Inappropriate according to operationally defined behavior. Correlation between observers was +.96.

Results

During the baseline period the child's behavior was considered disruptive an average of 54 percent of the time during the three structured activity periods and 14 percent during the four relatively unstructured periods. Statistical analysis indicated that during this stage a significantly higher percentage of Lisa's time was spent engaged in disruptive behavior during structured than unstructured activity periods (Table 1). Inappropriate behavior was significantly reduced (from an average of 31 percent to two percent of the time at school) and, concurrently, appropriate behavior increased after treatment began.

During the latter three stages of the study, no significant differences were indicated between percent of time engaged in disruptive behavior in structured or unstructured activity periods. Lisa spent an average of five percent of the school session engaged in disruptive behavior during Stage II (first treatment period), four percent in Stage III (Reversal), and two percent in Stage IV (second treatment period).

Discussion

The baseline data indicated the child spent an average of 31 percent of each school session engaged in inappropriate behavior. Inasmuch as Lisa's behaviors were highly disruptive, the teacher or her aides constantly attended to her either in response to or in anticipation of the behaviors. Although not recorded quantitatively, the experimenter noted close to a one to one ratio between manifestation of inappropriate behavior and adult attention. The teacher and aides were giving Lisa an inordinate amount of their time in comparison to that given the other 17 children. Although this attention appeared necessary to control her, their attending behavior was apparently reinforcing Lisa's inappropriate behavior.

Twenty-four time outs were administered during the first treatment period: eight in the first school session during the structured activity periods and 4, 2, 4, 2, and 4 respectively during the following five school sessions. The treatment procedure resulted in the suppression of much of the previous disruptive behavior. During the six days of Stage II, Lisa now spent an average of 95 percent of her time engaged in socially acceptable behavior. This gave the teacher more opportunity to praise and give her attention commensurate with that given the other children. After the disruptive behavior decreased, appropriate verbal and physical peer interactions were observed to increase appreciably. During the four unstructured activity periods duration of discrete disruptive behavior was observed to lessen in response to nonattention and possibly to the generalization of the effect of the time out procedure used in the structured periods.

In successfully administering the time

out procedure, the child's mother supported the teacher and her aides in their attempt to respond immediately and consistently to Lisa's behavior. After initial clarification of the behavior resulting in the time out, there was no further interaction despite Lisa's crying, screaming, and other behavior ploys. Verbal behavior observed in the hall during time out periods extinguished by the fourth day, and the child learned to tolerate the two-minute period without incident. Her delight upon reentering the classroom was apparent.

The overt improvement in Lisa's behavior had a pronounced positive effect on the teacher, the aides, the children, and her mother. They began to trust the acceptable behavior they observed. This change in attitude certainly added appreciably to the reinforcing value of the teacher's praise and the growing influence of the general classroom environment. The natural establishment of peer interaction and successful task completion as reinforcement contingencies may explain lack of regression during the three day reversal stage. In addition, the mere presence of adults in the classroom would exert some control over the disruptive behavior due to their association with punishment during the treatment stage. As suggested in a recent study by Scott, Burton & Yarrow (1967) the reversal stage does not represent a complete return to baseline conditions due to the intervening experience.

When treatment procedures were reestablished during the final stage of the study, Lisa's behavior showed further improvement. During the six school sessions 11 time outs were administered (3, 2, 3, 2, 1, 0), whereas the child now spent an average of 98 percent of the school session engaged in socially acceptable behavior.

As the level of disruptive behavior in the structured and unstructured periods was not equal during baseline, the data relative to the reduction of disruptive behavior following initiation of treatment cannot support a differential treatment effect in comparing time out with nonattention. In fact, the reduction of disruptive behavior should be viewed as resulting from the combined treatment procedures of time out, nonattention and social reinforcement. The systematic use of social reinforcement during all periods and the probable generalizations of the effects of time out and nonattention from one class period to another support this combined treatment interpretation.

Lisa's progress was followed up 30 days after completion of the initial study. Two observers following the data gathering procedures previously used reported that the child engaged in socially acceptable behaviors an average of 98 percent of the time during three successive school sessions. Treatment procedures were maintained with the exception that a teacher aide was designated daily to administer time out when needed. The teacher reported no problems in following treatment procedures and indicated satisfaction with Lisa's social development.

Problem behaviors of a nursery school child were significantly reduced and appropriate behaviors were increased through the systematic appli-

cation of a behavior modification program. During the three structured activity periods of the morning, the child was removed from the class following the disruptive behavior, while socially acceptable behavior was reinforced by teacher praise and attention. During the four relatively unstructured activity periods of the morning, inappropriate behavior was ignored, whereas the teacher's praise and attention again was made contingent upon the child making socially acceptable responses. The time out procedure was initiated following baseline observations which indicated that a significantly higher percentage of inappropriate behavior occurred during the three structured activity periods when this behavior was more disruptive to the class than during the four relatively unstructured periods of the morning when it was possible to ignore the child's deviant behavior. After treatment was initiated disruptive behavior was significantly reduced during all seven class periods. Data supported the conclusion that the combination of time out, nonattention and social reinforcement was effective in achieving more appropriate behavior.

References

Allen, K. Eileen and Harris, Florence R. Elimination of a child's excessive scratching by training the mother in reinforcement procedures. *Behav. Res. & Ther.,* 1966, **4**, 79–84.

Allen, K., Hart, B., Buell, J., Harris, F., and Wolf, M. Effects of social reinforcement on isolate behavior of a nursery school child. *Child Develpm.,* 1964, **35**, 511–518.

Baer, D. M. Escape and avoidance response of preschool children to two schedules of reinforcement withdrawal. *J. Exp. Anal. Behav.,* 1960, **3**, 151–59.

Baer, D. M. Effect of withdrawal of positive reinforcement on an extinguishing response in young children. *Child Develpm.,* 1961, **32**, 67–74.

Baer, D. M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. *J. Exp. Anal. Behav.,* 1962a, **5**, 525–529.

Baer, D. M. A technique of social reinforcement for the study of child behavior: Behavior avoiding reinforcement withdrawal. *Child Develpm.,* 1962b, **33**, 847–858.

Harris, F., Johnston, M., Kelley, C., and Wolf, N. Effects of positive social reinforcement on regressed crawling of a nursery school child. *J. Educ. Psychol.,* 1964, **55**, 35–41.

Hart, B., Allen, K., Buell, J., Harris, F., and Wolf, M. Effects of social reinforcement on operant crying. *J. Exp. Child. Psychol.,* 1964, **1**, 145–153.

Hewett, F. M. Teaching speech to an autistic child through operant conditioning. *Amer. J. Orthopsychiat.,* 1965, **35**, 927–936.

Johnston, M., Kelley, C., Harris, F., and Wolf, M. An application of reinforcement principles to development of motor skills of a young child. *Child Develpm.,* 1966, **37**, 379–388.

Patterson, G. R., and Brodsky, G. A Behaviour modification programme for a child with multiple problem behaviors. *J. Child Psychol. Psychiat.,* 1966, **7**, 277–295.

Risley, T., and Wolf, M. Establishing functional speech in echolalic children. *Behav. Res & Ther.,* 1967, **5**, 7388.

Scott, P. M., Burton, R. V., and Yarrow, M. R. Social reinforcement under natural conditions. *Child Develpm.,* 1967, **38**, 53–63.

Wahler, R. G., Winkel, G. H., Peterson, R. F., and Morrison, D. C. Mothers as behavior

therapists for their own children. *Behav. Res. & Ther.,* 1965, **3**, 113–124.

Wolf, M., Risley, T., Johnston, Margaret, Harris, Florence, and Allen, K. Eileen. Application of operant conditioning procedures to the behavior problems of an autistic child: a follow-up and extension. *Behav. Res. & Ther.,* 1967, **5**, 103–111.

Wolf, M., Risley, T., and Mees, H. Application of operant conditioning procedures to the behaviour problems of an autistic child. *Behav. Res. & Ther.,* 1964, **1**, 305–312

5.4 *The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct*

JACOB S. KOUNIN AND PAUL V. GUMP

Reprinted with slight abridgment from the *Journal of Educational Psychology,* 1961, **52,** 44–49, with permission of the authors and the American Psychological Association, Inc. Jacob S. Kounin is a developmental and social psychologist who has been at Wayne State University in Detroit since 1946. Paul V. Gump is a developmental and educational psychologist with the University of Kansas' Midwest Psychological Field Station in Oskaloosa, Kansas.

The usual way that teachers deal with classroom misconduct is not, of course, through behavior modification, but through direct intervention—scolding, punishment, shaming, and the like. We have seen from Gallimore's discussion that some of this treatment may have a reinforcing effect and thus lead to misbehavior becoming more firmly fixed in the learner's repertoire. Punishment and similar forms of harsh intervention can have effects on bystanders as well, as the paper by Kounin and Gump shows. Not only do they become more aggressive, but they seem less able to learn more positive forms of behavior. An overemphasis on what *not* to do seems to make children confused with respect to what *to* do.

This paper reports a portion of a research project pertaining to the management of children's behavior in classroom settings. Because so many teachers, especially beginners, verbalize considerable concern about discipline and control, we are focusing our current research in this area. While there is some relevant literature, such as that of Sheviakov and Redl (1944), based upon experience and insightfulness, we have been unable to locate any generalizations based upon data from research.

In a previous study by Kounin and Gump (1958) specimen-record types of observations were gathered of discipline incidents during the first week of kindergarten, focusing upon the triad of: a misbehaving child (target), a teacher doing something to stop the misbehavior, and a watching audience-child. Limiting our dependent variables to overt behavior we found that teachers' techniques of handling a misbehaving kindergarten child (target) did have different degrees of socializing success upon audience-children. A socializing success was defined as an observable reduction of overt misbehavior or an increase in conforming behavior (standing up "even straighter" in line). Control techniques high in clarity (defining the deviancy, specifying how to stop) were most successful. Control techniques high in firmness (standing closer to the misbehaving child, continuing to look at him until he stopped misbehaving) were successful only for audience-children who were themselves deviancy-oriented at the time. Control techniques high in roughness (anger, physical handling) were least successful and tended to be followed by behavior disruption (less involvement in work, overt signs of anxiety) rather than conformity on the part of audience-children. In terms of their effects, it is evident that roughness is a different dimension from firmness.

Since attitudes toward misconduct may also be affected by differences in control techniques we decided to study these as well. In an unpublished study of children at camp, P. Gump, B. Biddle, and J. Kounin found significant differences in attitudes toward camp misconduct held by campers who had effective counsellors as compared to campers who had ineffective counsellors. The counsellors, however, varied along many dimensions including punitiveness, goal-directedness, physical and psychological absenteeism, and others. The campers' attitudes toward misconduct also varied according to whether they were talking about camp, home, or school milieus. We decided, therefore, to limit the leadership dimension to punitiveness and the milieu to school.

It is postulated that aggression leads to counteraggression; it is further postulated that a punitive teacher has more power over her pupils than they have over her and that she blocks overt manifestation of pupils' aggression (observations in the classrooms of the punitive teachers selected for this study indicate that this second assumption is tenable). From these two postulations, we derive the following hypotheses:

1. That the school misconduct preoccupations of children with punitive teachers will contain more aggres-

sion than those of children with nonpunitive teachers.

2. That children with punitive teachers will be more conflicted about school misconduct than will children with nonpunitive teachers.
3. That the aggression needs and the conflict relating to misconduct hypothesized to exist among children with punitive teachers will detract from their concern with schoolunique values that are not directly related to misconduct.
4. The question may also be raised as to whether or not the amount of tension generated in the children with these particular punitive teachers is sufficiently great to reduce the rational qualities of their attitudes toward misconduct.

Method

Subjects

The subjects were 74 boys and 100 girls attending their first semester of the first grade in the public schools of a large city. They represented all the children from six home rooms of three schools, in from upper-lower to middle-middle socioeconomic neighborhoods.

Procedure

Overall school climate was controlled by selecting pairs of punitive vs. nonpunitive teachers from the same school. Three such pairs were obtained from three elementary schools.

The initial selection of punitive and nonpunitive teachers was obtained from the principal and assistant principal. Following this the classes were observed by both principal investigators. At approximately a week later the teachers were further rated by a supervisor of student teachers who visited each class twice.

The raters checked along a continuum from Extremely Punitive (threatens children with consequences that really hurt; makes threats that imply sharp dislike, real willingness to harm child; ever-readiness to punish) to Nonpunitive (does not punish and does not threaten). A punitive vs. nonpunitive pair of teachers was used for the study only when all five persons agreed on their dichotomizations. All the teachers were rated as having good organization, well-behaved classes, and as achieving the learning objectives for their grade. Eighty-four of the children were in classes with punitive teachers and 90 children were in classes with nonpunitive teachers.

The children were interviewed individually during the third month of attendance at school. The interview consisted of the questions: "What is the worst thing a child can do at school?" and, following the reply, "Why is that so bad?" Identical questions were asked regarding home as the milieu for misbehavior.

Coding the replies

The misconducts mentioned by the children were coded for content and for certain qualities or dimensions.

The content code (obtained from the question of "What is the worst thing to do?") contained two parts: the misconducts and the explanations given

for why these were bad. The misconduct included: the act type (physical or psychological assaults, noncompliance, etc.) and the object of the misconduct (parents or teachers, other children, institutional laws or custom, etc.).

The code for the explanation of misconduct was designed to answer three questions: Who is involved in the consequence (the child himself, parent or teacher, a peer, etc.)? What kind of sufferings result to others from the misconduct (physical pain, achievement loss, property loss, etc.)? What kinds of retributions occur to the misbehaver (work imposal, character loss, physical punishment, etc.)?

Results and Conclusions

Children probably answer the question of "What's the worst thing a child can do in school?" with a report of acts that reflect their preoccupations. It is not likely that our subjects' answers would have been the same if they were presented with a forced-choice of alternative acts. Given a choice, most children would probably rate "stabbing someone" as more serious than "talking in class." If the misconducts the children talked about are taken to represent tension systems and preoccupations, we may infer from these the comparative impact of punitive and nonpunitive teachers.

In a concurrent study of children's attitudes toward misconduct (201 boys and 214 girls in the first grade of six public schools representing a range of socioeconomic backgrounds), Gump and Kounin (1959) found both sex differences and differences between home and school milieus. For example, home misconducts included more breaking of objects while school misconducts included more rule violations; parents suffered more than teachers in consequences but teachers retributed more frequently. However, parents were reported as retributing with more corporal punishment with more severe punishment than teachers. There were also differences in responses of boys and girls, especially in school. For example, girls reported "talking" as a school misconduct eight times more frequently than boys, whereas boys reported physical assaults on peers in school more frequently than did girls.

Consequently, the comparison of the responses of children with punitive and nonpunitive teachers was made separately for sexes and also for home and school milieus. However, on *all* comparisons of school responses the direction of differences between children with punitive and nonpunitive teachers was the same for both boys and girls. The report of results, therefore, combines both boys and girls. Insofar as the differences between children with punitive and nonpunitive teachers are concerned, only 2 of the 48 comparisons of home responses were statistically significant: home misconducts of the children with punitive teachers were rated as more serious ($p < .05$, for girls only) and retributions to the subject were more serious ($p < .02$, for boys only). It is uncertain whether these represent some spillover of the influence of punitive teachers onto attitudes toward home misconducts, or whether they are chance differences for the number of comparisons made.

The results to be reported here, then, refer to boys and girls combined and to school misconducts only. Intercoder reliabilities ranged from 73-95% agreement, with a median of 90. The *p* levels of differences are based on the χ^2 test. In the case of dimensions, such as "seriousness," the results were dichotomized into a High and a Low based upon as equal a break as was possible and resulting χ^2's were based upon 2 × 2 tables. In the case of categories such as act-types falling into categories of: rule violations, physical assaults on children, property damages, or nonconformance with adults, the χ^2 was computed for as many cells as there were categories. At times, when one particular category was of interest, a 2 × 2 table was constructed with that category versus "all the rest," providing that the overall table showed statistical significance.

Following are definitions of the codes used which are not self-explanatory both for the misconducts and for the subjects' explanations for the wrongness of the act. These codes appear in Table 1.

I

A. Physical assaults include all physical attacks on other persons (pushing, hitting).

TABLE 1. A Comparison of Attitudes toward School Misconducts Held by Children with Punitive and Nonpunitive First Grade Teachers (N = 176)

Misconducts and Explanations	% of PU [a]	% of NPu [a]
I. Content and quality of the misconducts:		
A. Physical assaults on others	38	17
B. Milieu-serious misconducts	89	63
C. Coder-serious misconducts	48	27
D. Abstract misconducts	27	52
II. Content and quality of the explanations:		
A. Peers as objects of consequences	94	61
B. Physical damage to objects of consequences	60	23
C. Serious harm to others	45	18
D. Reality-centered retributions	21	48
E. "Reflexive justifications" as explanations	11	26
III. Role of self in misconducts:		
A. Ego-alien misconducts	26	11
B. Premeditated misconducts	29	15
IV. Aggression:		
A. Overall aggression ("blood and guts")	49	24
V. Concern with school-unique objectives:		
A. Learning and achievement losses	20	43
B. Institutional law violations	49	62

Note. All differences in percentages are significant at the .05 level or beyond.
[a] Pu stands for those children who have punitive teachers; NPu refers to those children who have nonpunitive teachers.

B. Milieu-seriousness refers to the length to which the milieu would go to prevent such an act. The school would practically ignore "scratching head," would mildly frown at "whispering," and would go to any length to stop burning down buildings.

C. Coder-seriousness refers to the general immorality or danger in the misconduct considered from the point of view of the overall culture. (The only frame of reference which produced high intercoder agreement was when the coders took the position of an understanding Reformed rabbi or a Unitarian minister. Taking an unspecified role, or that of either a parent or a teacher produced low intercoder agreement.) "Studying spelling lessons at the wrong time" is morally trivial while "maiming someone" is morally very serious.

D. Abstractness of misconduct refers to the size of coverage. It may range from a unique, "one time" misbehavior, such as, "cut a climbing rope in gym," to an abstract one, such as, "be mean to other people."

II

A. A central adult is the responsible leader: teacher at school.

B. A psychological loss to another is exemplified by "It would make her worry."

C. Seriousness of consequences to others ranges from trivial harm, such as, "She'd be annoyed" to serious ones, such as, "He'd die."

D. A reality-centered retribution (this is scored only when the perpetrator himself suffers in the consequence) is coded when the consequence of a misconduct follows naturally from the act-type, such as: "not study because you'll get behind in your work." This contrasts with the response in which the connection between act and consequence is dependent upon a personal intervention of another, such as: "not study because teacher will make you stand in the corner."

E. "Reflexive justification" was coded when the child gave no consequence for either himself or others in his explanation of why the act was bad. When he said the act is bad because "It's not nice" or "It's bad" it was called a reflexive justification.

III

A. On ego-acceptability, we sought to determine the degree to which the respondent could see himself as the perpetrator of the misconduct. In an ego-alien act, the respondent expresses abhorrence, such as: "It's dirty to hit little kids who didn't do nothing to you." An ego-attractive act is one in which the child indicates its seductive quality for him, such as: "Tell off a teacher—boy, I'd like to do that."

B. On the premeditation category, we sought to learn the extent to which the child sought to do wrong. If premeditated, the child plans the act or intends the consequences ahead of time, such as: "Put thumb tacks on teacher's chair when she is out." If intentional, the child accepts his part in the wrongdoing but does not plan it, such as, "talk during a lesson."

IV

Aggression ("blood and guts") refers to the amount of aggression the respondent expresses in his misconducts and consequences. "Play in the storage bin because somebody might get hurt" expresses less aggression than "Play in the storage bin because you might push a kid off and there could be a sharp rock down there and he could hit his head against it and crash open his skull and he would bleed and his brains would fall out and he'd die."

V

A. A learning or achievement loss is coded when interference with learning is the misconduct or the explanation, such as: "It's bad to make noise because somebody could make a mistake in his work," or ". . . because then he couldn't read good.

B. An institutional law violation is a violation of the rules of the school such as: "talk when you're supposed to study," "not take your seat when the bell rings."

The results presented in Table 1 may be summarized around the three hypotheses and the one question raised in the introduction:

Punitive teachers will create or activate more aggression-tension than will nonpunitive teachers. This is strongly supported by the data. The children who have punitive teachers have more sheer aggression in their sins and consequences, they give both more milieu-serious and more coder-serious misconducts, their targets suffer more harm, they give more physical assaults as act-types, and their targets suffer more physical harm. The targets of children with nonpunitive teachers are more inclined to suffer psychological losses as consequences. As an example of the results: of 84 respondents with punitive teachers, 31 gave physical assaults on other children and 40 mention school rule violations; while of 90 children with nonpunitive teachers, 15 talk about physical assaults and 56 about rule violations. (The remainder of the act-types are nonconformances and "miscellaneous.")

Children with punitive teachers will be more unsettled and conflicted about misbehavior in school. This hypothesis is supported by the findings related to the role of self in misbehavior. The children from nonpunitive teachers give misconducts in which their own role is intentional whereas children from punitive teachers give both premeditated and ego-alien misconducts. We may say that children with punitive teachers express more abhorrence for the misdeeds which they have selected and yet select misdeeds which require "malice and forethought."

Punitiveness of teachers will detract from children's concern with school-unique values. This hypothesis is supported. Children from punitive teachers talk more about physical attacks on peers — misbehavior by no means unique to the classroom setting. Children with nonpunitive teachers talk more about learning, achievement losses, and violations of school-unique values and rules.

Do children from nonpunitive teach-

ers show more rational qualities in their responses? The answer to this question is not clear. Fairly direct attempts to measure this—codes for milieu likelihood of misconducts, for likelihood of consequences, and for appropriateness of consequences to the misconduct—did *not* show significant differences between the two groups. On the other hand, children with punitive teachers gave fewer abstract misconducts which result, in our camp study, was negatively correlated with age. But these same children also gave fewer reflexive justifications which result was positively correlated with chronological age. One interpretation of the findings that children of punitive teachers gave fewer abstract misconducts and fewer reflexive justifications is to regard these as indications of the unsettled and conflicted state of the attitudes regarding misconduct held by children with punitive teachers. When a child is inclined to misbehave but fears to, then a concrete act occurs to him—"hit George in the mouth"; when he is not pressed by his needs to misbehave, then an abstraction occurs to him—" be mean to people." Similarly, when he expresses this verbal act, a real consequence occurs—the target gets hurt or the perpetrator suffers a consequence; when he is not preoccupied with wrongdoing then a reflexive justification occurs to him—"it's not nice." A reflexive justification at this age may not be a primitive reply but a reflection of a settled issue: "You just don't do this because it's not nice."

Another interpretation is to regard the greater use of reflexive justification by the children with nonpunitive teachers as evidence of their greater trust and faith in school, i.e., of their internalization of school values more than children with punitive teachers. Inspection of the data showed the reflexive justification was used predominantly in connection with rule violations (talking, running in halls, not taking seat, and the like). These misconducts are milieu-inconvenient which are disturbing to the milieu but which are without direct harm to either the actor or to others and do not violate an important moral code. Such misconducts to the first grade child have no real explanation except that "they're bad because they say so." As such, they express a sort of naive faith and trust in the rightness of what the teacher says.

Summary

Three pairs of punitive vs. nonpunitive first grade teachers were selected from three elementary schools. The 174 children in these teachers' classrooms were individually interviewed about what they thought was "the worst thing to do in school" together with their explanations of why these misconducts were bad. Regarding their responses as expressions of their preoccupations it was concluded that, as compared with children who have nonpunitive teachers, children who have punitive teachers: manifest more aggression in their misconducts, are more unsettled and conflicted about misconduct in school, are less concerned with learning and school-unique values, show some, but not consistent, indication of a reduction in rationality pertaining to school misconduct. A theory that children with

punitive teachers develop less trust of school than do children with nonpunitive teachers was also presented to explain some of the findings.

References

Gump, P. V., and Kounin, J. S. Sex and milieu influences in children's concepts of misconduct. *Amer. Psychologist,* 1959, **14**, 346. (Abstract)

Kounin, J. S., and Gump, P. V. The ripple effect in discipline. *Elem. Sch. J.,* 1958, **59**, 158–162.

Sheviakov, G., and Redl, F. Discipline for today's children and youth. Washington, D. C.: National Education Association, 1944.

5.5 *Changes in Pupil Attitudes During the School Year*

NED A. FLANDERS, BETTY M. MORRISON, AND E. LELAND BRODE

Reprinted from the *Journal of Educational Psychology,* 1968, **50**, 334–338, with permission of the authors and the American Psychological Association, Inc. Ned A. Flanders is with the Far West Regional Laboratory in Berkeley, California, and is well known for his work on the social psychology of the classroom. Betty M. Morrison and E. Leland Brode are on the faculties of Western Reserve University and Northern Illinois University, respectively, and have done research in interaction analysis.

Although teachers have a major influence on the psychological aspects of what goes on in the classroom, we should keep in mind that students, too, are a source of influence. The research paper by Flanders, Morrison, and Brode deals with a dimension that is often overlooked both by teachers and educational psychologists: the shift in students' attitudes toward their teachers. Their findings that attitudes become less favorable as the school year progresses is significant, for the shift both contributes to and is a result of failures in the teaching-learning process. Teachers tend to take student attitudes for granted. If they are going to develop a better understanding of classroom dynamics, they should keep in mind that student attitudes are likely to worsen over time. Deteriorating attitudes can, furthermore, interfere with teacher-student communication and, eventually, with learning. The authors do not tell us what we can do about this problem, but they have done us a service in bringing it to our attention.

Are pupils most optimistic about schoolwork as the school year begins? Does this optimism erode as the school year progresses? Are there particular patterns of teacher behavior which appear when there is less erosion of optimistic pupil attitudes? This article will attempt to answer these questions, not with complete, unequivocal answers, but with some suggestions based on two separate studies.

In a 1960–61 Minnesota study (Flanders, 1963), fairly conclusive evidence was collected indicating that over 3,000 students in two junior high schools scored highest on an attitude inventory assessing positive perceptions of their teachers and their schoolwork in October, only to have a statistically significant decrease in the scores of a January readministration of the same inventory. A follow-up administration was about the same as January, significantly lower than the October scores.

The 1960 attitude inventory consisted of 59 items roughly divided into four subscales on the basis of content: (*a*) teacher attractiveness, which included such items as, "I would like to have this same teacher next year," and, "This is the best teacher I ever had," (*b*) fairness of rewards and punishments, which included such items as, "This teacher punishes me for things I didn't do," and, "This teacher punishes the whole class when he (she) can't find out who did something," (*c*) teacher competence, which included such items as, "Our teacher is very good at explaining things clearly," and, "It is easy to fool this teacher," and (*d*) interest in schoolwork, including such items as, "This teacher makes everything seem interesting and important," and, "Most of us get pretty bored in this class." The response to each item was on a 5-point scale from strongly disagree to strongly agree. All items were keyed so that a higher score represented more positive attitudes and perceptions.

The mean score of the October administration of the attitude inventory was 217. The means of the January and May administrations were 204 and 205, respectively—both significantly lower ($p < .01$) than the October administration. These data were collected in so-called academic classes in Grades 7, 8, and 9, excluding such subjects as physical education, music, home economics, and shop.

The results seem quite clear. There is a significant reduction in the average scores of positive pupil perceptions between October and January of the school year.

The Present Study

During the 1964–65 school year a Michigan Student Questionnaire (MSQ) was administered to 101 sixth-grade classes in 15 school districts near Ann Arbor. Thirty classes were selected for further study from the October distribution to include the top 10, the bottom 10, and 10 near the average of the 101 classes. The test was readministered in January and again in May in these 30 classes, each administration involving more than 800 pupils, and the sample can be considered representative of over 3,000 pupils who were in the the larger population.

The MSQ was essentially the same

inventory used in 1960–61 except that the items had been simplified in an effort to adjust the vocabulary to the reading skills of sixth-grade pupils. A factor analysis of the MSQ indicated that the most important factor was teacher attractiveness, with additional factors of teacher competence, teacher fairness, and lack of pupil anxiety forming a combination which was less important than the first factor.

The means for each of the 30 classes are shown in Table 1, arranged in terms of the top, middle, and bottom 10 classes on the October administration.

The 1964–65 results were nearly identical to the 1960–61 results. There was a significant drop in average scores of positive pupil attitudes during the first 4 months of the school year. The mean score for the October administration was 178.2 with a standard deviation of 26.52. The January administration had a mean of 172.2 and a standard deviation of 31.13; and the May administration had a mean of 170.6 and a standard deviation of 30.60.

The rest of this article will discuss the various factors that might be related to the observed change in attitude, one administration compared to the next.

Factors not associated with change in attitude

Simple regression is not an adequate explanation of these changes. While it is true that three low classes (Classes 21, 24, and 27) showed the highest positive changes, it is also true that three other low classes (Classes 22, 23, and 28) showed large decreases. The average loss (October to May) of the bottom 10 classes was 6.2, compared to 7.6 for the total group. Furthermore,

TABLE 1. Means of 30 Classes On the 1964–65 Michigan Student Questionnaire

Class	Administration			Class	Administration		
	Oct.	Jan.	May		Oct.	Jan.	May
1	204.9	194.7	194.2	16	178.8	178.0	176.1
2	201.4	200.0	195.4	17	177.9	185.4	178.2
3	200.3	204.9	200.0	18	176.2	166.6	169.3
4	199.6	194.7	192.3	19	175.4	169.2	167.0
5	197.5	195.2	195.2	20	173.8	170.2	166.5
6	197.0	193.5	189.9	21	166.4	173.1	175.0
7	193.6	190.6	187.9	22	166.2	147.7	150.4
8	192.0	185.7	175.9	23	165.1	151.3	143.8
9	191.6	178.9	186.5	24	162.7	174.7	173.9
10	190.3	178.3	176.3	25	162.7	159.8	160.8
11	180.9	163.7	156.3	26	158.1	149.4	147.2
12	179.2	180.3	176.2	27	158.1	154.0	161.2
13	178.8	173.3	164.5	28	157.2	145.5	139.2
14	176.8	171.4	155.7	29	156.6	137.3	147.7
15	175.2	151.9	155.7	30	149.9	137.2	142.2

the top 10 classes do not show uniform loss but instead are symmetrically distributed about an average loss of 7.5.

The correlations between administrations, based on individual scores, were positive and fairly high: for October to January, $r = .704$; for January to May, $r = .812$; and October to May, $r = .655$. The correlation between October and May, based on 30 class averages, is higher; $r = .876$. The correlations present a picture of a fairly stable response pattern both within and between classes.

There is the possibility that change in positive pupil attitudes might be associated with the average class IQ, socioeconomic status, or the percentage of A and B letter grades assigned to the pupils by the teacher. Table 2 shows such data for the nine classes which had high change losses and the seven classes with the least amount of change. The mean IQ for the high-change group was 113.5, while it was 112.3 for the low-change group. Here the IQ scores used were those based on school records and probably involved different published tests. The median socioeconomic rating for the high-change group was 71; a median rating of 70 was obtained for the low-change group. Here a rating on the National Opinion Research Center scale (Reiss, 1961) was made of the wage earner's occupation as reported by the teacher. The mean percentage of A and B letter grades for the high-change group was 56.5 and for the low-change group, 64.5. While this last difference is consistent with a theory that change involving loss of positive attitudes is associated with receiving lower grades, a z test between independent proportions yielded a value of 1.66 which was not high enough to reject the null hypothesis at the .05 level of significance. All of these data suggest that changes in class attitudes are not significantly associated with

TABLE 2. Comparisons Between High- and Low-Attitude Change Classes

High change				Low change			
Class no.	IQ	SES rating	% of A & B grades	Class no.	IQ	SES rating	% of A & B grades
11	121.6	81	75	3	118.1	65	91
23	104.2	67	33	17	114.1	78	55
14	111.7	71	64	25	104.1	69	48
15	118.5	67	52	5	113.1	70	66
28	109.8	74	41	16	120.7	71	81
8	116.8	78	57	12	100.0	63	50
22	110.3	67	63	27	116.1	83	62
13	116.0	83	80				
10	112.7	70	42				

Note. SES = socioeconomic status.

average IQ, socioeconomic status, or grades given by the teacher.

Two factors associated with change in attitude

In another study, Morrison (1966) has shown that Rotter's notion of "externality" and "internality" (Rotter, Seeman, & Liversant, 1962) can be assessed among sixth-grade pupils. By externality is meant the tendency of a pupil to believe that his successes and failures are caused by forces beyond his control. By internality is meant the tendency to believe that successes and failures are self-determined and products of one's own behavior. External children, according to Morrison's conception, would be more likely to associate the good and bad outcomes of classroom learning activities with the teacher who is a powerful source of influence. Internal children, on the other hand, would see themselves as more closely associated with the good and bad characteristics of learning outcomes.

A test of internality-externality was administered to all the pupils in the 30 classes during the January administration of tests. The test consisted of 26 items, each containing two statements, and the pupils responded by marking the statement in each item which they believed was more often true. Typical items were: (*a*) "If you study you will do well on a test," or (*b*) "People who score the highest on a test are lucky." —(*a*) "Most of the time children get the respect they deserve from others," or (*b*) "Many times a child can try hard and no one will pay attention to him." —(*a*) "Usually other people choose me for a friend," or (*b*) "Usually I choose my own friends."—(*a*) "Children get into trouble because their parents punish them too much," or (*b*) "The trouble with most children is that their parents are too easy with them."

Each item was scored 1 if the internal response was chosen and 2 if the external response was selected, giving a possible range from 26 to 52 for the total scores. The actual scores ranged from 26 to 49. Students in the lower third of the distribution (raw scores of 31 and below) were defined as internals and those in the upper third (raw scores of 36 and above) were defined as externals.

In addition to these tests each of the 30 classrooms was visited by an observer trained to code verbal communication into the 10 categories of interaction analysis developed by Flanders (1965). More than six visits were made to each class and more than 7,000 tallies were recorded by observers. The main results of interaction analysis will be reported elsewhere; the interest for the moment is in the incidence of praise and encouragement expressed by the teachers during these visits. The occurrence of this type of teacher statement varied from low of 2% to a high of 2.1% of all tallies recorded by the observer. The problems of reliability among observers and the representativeness of the interaction sampled are too complex to be discussed here. It can be said, however, that the relative objectivity of the observation data, or lack of it, would affect the data from all classes equally and cannot account for

any of the differences about to be discussed.

It was hypothesized in this study that:

1. External children have a greater negative shift in attitude than do internal children.
2. The classes of low-praise teachers have a greater negative shift in attitude than do the classes of high-praise teachers.
3. The attitudes of external children are more affected by the praise and encouragement of the teacher than are the attitudes of internal children.

To test these hypotheses a two-way analysis of covariance in the case of unequal or disproportionate numbers of observations in the subclasses was performed using the third attitude inventory scores as the dependent variable and adjusting with the scores from the first administration. Table 3 includes the October means, the May means, and May means adjusted for the initial attitudes, and the change means. These are arranged in subgroups of internal and external pupils and pupils with high-praise and low-praise teachers.

The slope of the regression line was .845, and the analysis of covariance yielded an error mean square of 520.7. The main effect for internal versus external pupils produced a mean square of 18,331 and the resulting F ratio is significant at well beyond the .01 level ($F = 35.20$, $df = 1/473$). The main effect for pupils of high-praise versus low-praise teachers produced a mean square of 7,128, also resulting in an F ratio which is significant at well beyond the .01 level ($F = 13.69$, $df = 1/473$). However, the interaction of pupil types and teacher styles resulted in a mean square of only 90, which is not significant ($F = .17$).

TABLE 3. Initial, Final, and Adjusted Means of Attitude Scores from Analysis of Covariance by Pupil Type and Teacher Style

Teacher style	Pupils		
	Internal	External	All
High-praise			
Initial *M*	190.6	171.0	183.9
Final *M*	187.3	159.2	177.7
Adjusted *M*	163.7	152.1	159.7
Change *M*	—3.3	—11.8	—6.2
Low-praise			
Initial *M*	178.7	121.5	143.9
Final *M*	173.1	111.3	135.5
Adjusted *M*	159.5	146.0	151.3
Change *M*	—5.6	—10.2	—8.4
All			
Initial *M*	185.8	137.8	162.6
Final *M*	181.6	127.1	155.2
Adjusted *M*	162.0	148.0	—
Change *M*	—4.2	—10.7	—

These results indicate that not only did external pupils have less positive attitudes than did internal pupils early in the school year, but when the May scores are adjusted by the October scores, it is apparent that external pupils experienced significantly greater declines in their attitudes than did internal pupils. Also, pupils with low-praise teachers showed greater losses in positive attitudes during the year than did pupils with high-praise teachers. However, there was no evidence that the attitudes of external children were more affected by praise or lack of praise on the part of the teacher

than were the attitudes of the internal children.

Discussion and Conclusions

In two separate projects the attitude inventory scores indicate that positive perceptions of pupils toward their teacher and their class activities decrease sometime during the first 4 months of the school year. In the second project these changes were shown to be unrelated to IQ, performance grades assigned by the teacher, and the socioeconomic ratings of the father's occupation. Two hypotheses about changes in attitude were supported. First, external pupils experience a greater loss of positive attitudes toward school than do internal pupils. Second, in classrooms of teachers who provide less praise and encouragement there is greater loss of positive attitudes than in classrooms of teachers who provide more praise and encouragement. The third hypothesis was not supported since the interaction effects in Table 3 are not significant.

One inference to be drawn is that the type of youngster who is more dependent on external influences seems to be more likely to suffer a loss of positive expectations than is the one who is more dependent on internal influences. In addition, pupil attitudes toward the teacher and the learning activities seem to be related to teacher behavior. Whether this difference in pupil attitudes is the result of the different teacher behaviors, or the different amounts of teacher praise and encouragement are the result of the pupils being more or less deserving of that praise, is not clear from the evidence of the present study. The absence of a significant difference between high-change and low-change classes with respect to the percentage of A and B grades given by the teacher (Table 2) would indicate that the pupils' performance was not the deciding factor. Also, previous studies (Flanders, 1963, 1965) have indicated that teacher behavior is the more dominant factor and that differences in such patterns of teacher influence tend to be greater between different teachers than between different situations for the same teacher.

In this sample, differences among the pupils had a greater effect than the presence or absence of a small amount of teacher praise and encouragement. Future studies of the erosion of positive pupil attitudes may wish to take into account other differences among pupils as well as differences in teacher behavior.

This study did not provide direct evidence concerning two opposite hypotheses about the erosion of positive pupil attitudes. One theory is that the pupils become disenchanted with the teacher during the first few months of the school year. A second theory, based on the assumption that the October scores are inflated or too high, is that as the pupils learn to trust their teacher they do not overestimate their ratings as they felt compelled to do with a strange teacher. Without going into detail, the authors tend toward the first of these two theories, primarily because the teacher's behavior is the predominant influence in the typical classroom; but much more evidence

will be required before any conclusions can be reached.

Meantime, lack of loss of positive attitudes may be the mark of a good match between teaching behavior and particular attributes among pupils. Apparently, in most classrooms such a match does not exist.

Summary

Data from 2 separate studies indicate a significant loss in positive attitudes of pupils toward their teachers and schoolwork during the school year. In the present study of 820 6th-grade pupils in 30 classrooms it was shown that this erosion of positive attitudes is not related to pupils' IQ, socioeconomic status, or percentage of A and B letter grades assigned by the teacher, but is related to the "externality" or "internality" of the pupils and to the teachers' verbal classroom behavior. Greater losses in attitudes occurred among external than among internal pupils and among pupils whose teachers exhibited a lower incidence of praise and encouragement than among those whose teachers exhibited a higher incidence of such behaviors.

References

Flanders, N. A. Helping teachers change their behavior. Terminal report, National Educational Defense Act, Title VII project, 1963.

Flanders, N. A. *Teacher Influence, Pupil Attitudes, and Achievement.* United States Office of Education Cooperative Research Monograph No. 12. Washington, D. C.: United States Government Printing Office, 1965.

Morrison, B. M. The reactions of external and internal pupils to patterns of teacher behavior. Unpublished doctoral dissertation, University of Michigan, 1966.

Reiss, A. J., Jr. *Occupations and Social Status.* New York: Free Press of Glencoe, 1961.

Rotter, J. B., Seeman, M., and Liversant, S. Internal versus external control of reinforcements: A major variable in behavior theory. In N. F. Washburne (Ed.), *Decisions, Values, and Groups.* Vol. 2. London: Pergamon Press, 1962.

5.6 *What Are We Rewarding?*

JAMES A. CONWAY

Reprinted from the *Phi Delta Kappan,* 1969, **61,** 187–189, with the permission of the author and Phi Delta Kappa, Inc. James A. Conway is a member of the education faculty of the State University of New York at Buffalo.

As we interact with others, many of the dimensions of our behavior are beyond ordinary awareness. Gallimore has shown how we inadvertently reinforce misbehavior, and in this paper Conway suggests that the students to which we give most attention are likely to be those that are more dogmatic and closed-minded than the other members of the class. Inasmuch as teacher attention is a form of reinforcement or reward, his finding is somewhat unsettling, for most teachers probably believe that their attention is directed primarily at the more receptive, open-minded students.

To most teachers the following remarks will sound familiar. In the process of reading examination papers (or compositions, or term papers) the instructor says:

I can't understand how John could have written such a poor paper. He's a good student, I'm sure of it. He's alert in class, asks questions, participates, a provocative student—but this paper just doesn't sound like him.

And here's Don's paper, another surprise! He really got to the point; his answers are clear and concise, surprisingly perceptive. I wonder why he never says anything in class? Come to think of it, I can't even picture which of my students Don is.

Now how am I going to grade these papers? Do I mark John down for this paper when I know he's a good student? *Maybe he just had a bad day! Should I allow each student to select one test grade to be eliminated? Should I be putting so much weight on these written responses?*

Whether the situation is familiar or not, it is one that exists. Ebel states, "Often a student's mark has been influenced by the pleasantness of his manner, his willingness to participate in class discussion, his skill in expressing ideas orally or in writing, or his success in building an image of himself as an eager, capable student."[1] Palmer indicates that marks are used as rewards or punishments for certain *kinds of behavior* manifested in the classroom.[2] The fact that grades may be tempered by classroom behavior is, perhaps, disturbing, but it is only half of the picture. My purpose here is to look at what we may be rewarding in such cases.

An incident that occurred while I was studying for my doctorate prompted me to pursue further the situation described above. I had administered a personality test to some 700 students enrolled in undergraduate education courses. The test was Rokeach's Dogmatism Scale, which purports to measure the degree to which a person's system of beliefs is open or closed to new or differing beliefs.[3] Rokeach points out that the basic characteristic that defines the extent to which a person's system is open or closed is:

. . . the extent to which the person can receive, evaluate, and act on relevant information received from the outside on its own intrinsic merits, unencumbered by irrelevant factors in the situation arising from within the person or from the outside.[4]

From the pool of students tested, I selected a sample of extremes; that is, students predominantly open-minded and students predominantly closed-minded. To contact the subjects, I approached professors teaching the undergraduate education courses and asked them if I could talk to particular students. I would begin reading off the names. After no more than three or four, professors would invariably smile and say something like, "Yes, I can see why you are selecting those students—they are all sharp, perceptive." At this point I would check scores, finding, almost without exception, that these students were predominantly closed-minded as measured by the test. When I continued, reading names of

the open-minded subjects, I found the professors having difficulty recalling them; in some cases they couldn't.

I was puzzled. Why were the closed-minded students easily identified and the open-minded students for all practical purposes anonymous? Rokeach noted that because the teacher only sees the student in a teacher-student context, the judgment may be quite unreliable.[5] This is understandable, but it doesn't explain the discrepancy between those who are recognized and those who are nameless.

Other investigators have reported related phenomena. Getzels and Jackson, as well as Alexander, indicate that conforming students seem to be rewarded with higher grades.[6] Dressel and Lehmann found that dogmatic and stereotypic students tend to receive higher grades from their instructors than the general academic aptitude of such students may warrant.[7] These and other studies suggest that the factor of dogmatism or open- and closed-mindedness may be directly involved in the biases that influence marks given as a result of teacher-pupil relationships in the learning situation.

Rokeach also argues for this position. He indicates that the dogmatic person is typically sensitive to the presence of authority in a social environment. In such an environment the closed-minded person is likely to adopt a respectful and acquiescent, yet at the same time enlightened and objective facade. He says further that the open-minded person's "equalitarian appreciation of his academic environment . . . may lead him to a greater willingness to venture sincere opinions which sometimes challenge or amend those made by his professors."[8] While Rokeach's position may be in agreement with the studies mentioned above, it is not consistent with my experience as the "questioning graduate student." Rokeach's "acquiescent student" was the individual I found the professors recognizing; the student who, according to Rokeach, *may* venture "sincere opinions and challenge professorial positions" was the unknown.

This apparent inconsistency prompted another investigation, made with John R. Dettre.[9] Its purpose was to determine the extent to which congruence (or noncongruence) of belief systems of students and of teachers influenced the final marks given in courses in teacher education on the undergraduate level. Twenty-six faculty members and 792 students at the State University College at Buffalo, New York, were administered Rokeach's Dogmatism Scale and also an adaptation of the life goals used by Getzels and Jackson.[10] Midway through the semester faculty members were asked to rate each of their students as being "above average," "average," or "below average." The ratings were to be general impressions of the student without reference to grade books or other sources. At the end of the semester final grades were recorded for each of the student subjects.

In order to magnify any differences, students and faculty whose scores on the Dogmatism Scale were in the top and bottom quarters were selected for analysis. Groups of closed-minded and open-minded students who had a closed-minded instructor, and also

closed-minded and open-minded students who had an open-minded instructor, were formed for chi square analyses of grade distributions. Similar groups were also formed of high and low student-faculty goal agreements, as well as groups stemming from the interaction of belief systems with goal agreement as related to final grades. The last analyses were of the mid-semester estimates made by open- and closed-minded faculty members of their respective open- and closed-minded students.

Starting with the mid-semester estimates first, it was found (at the .01 confidence level) that closed- and open-minded faculty do rank their closed- and open-minded students differently. Closed and open faculty, considered together, assigned significantly more above-average ratings to the closed students than to the open students. When considered separately, it was found that the closed faculty ranked their closed students higher than their open students, but that the open faculty did not differ in their ranking of open and closed students. Agreement or lack of agreement on goals between students and faculty did not seem to influence the rank that students received in this midsemester rating process.

When the distributions of final grades were analyzed, a slightly different picture emerged. Neither the variable of open- and closed-mindedness nor the variable of high and low goal agreement seemed to be related to the distribution of final grades among the students. However, when the interaction of the two variables was examined in relation to grade distributions, then a number of significant results appear. The one outcome that seems to be of particular interest was that those closed-minded students who were in *low* agreement with their instructors received a significantly higher proportion of A's and B's than any other category of students, including those closed-minded students whose goal orientation was in *high* agreement with their particular faculty member. What is it about the variable of dogmatism combined with low goal agreement that influences the reported outcome? While no definitive conclusions may be stated at this time, I would like to consider the question and offer a tentative explanation.

Some of the topics treated in undergraduate education courses deal with such factors as "child-rearing practices," "permissiveness," "conditioning," "teaching machines," "authoritarianism," and so forth. It is possible that some of the areas of study may be in conflict with the beliefs of some students. A portion of the closed-minded students may find that their system of beliefs is threatened as they treat such topics in the classroom group. Rokeach points out that the closed-minded person will try to protect his threatened belief system by warding off the threats.[11] Conway found that, when all the members of a group were predominantly closed-minded, the individuals warded off threats by refusing to contribute to the situation. By refusing to discuss a concept or topic, the closed-minded members avoided having to take a stand on the belief in question. However, in mixed groups,

which are more likely in the classroom situation, the closed-minded person cannot deter a conflicting belief through silence; instead, he must externalize irrelevant internal pressures, vocalizing or verbalizing frequently so as to direct the discussion away from the emerging threat.[12]

It may be that *low* agreement with the goal orientation of the professor is an indication of conflicting beliefs. When the low agreement is combined with a closed system of beliefs, then the situation is established for the argumentative verbalizations of the closed-minded person. It is almost a certainty that in almost all of the undergraduate education courses there is an attempt to foster the freedom to explore concepts and listen to opposing points of view; as such, the opportunity is created for exploitation by the closed-minded person. He is, in effect, *expected* to defend his system of beliefs and prevent the alteration of that system.

On the other hand, the open-minded person, and the closed-minded person in high agreement with professorial goals, may react little if at all in the class situation. What reason is there to be vocal when you are in agreement with the beliefs stated? In those cases where the open-minded person is in *low* agreement with the instructor he may either internalize the new beliefs and integrate them into his open system, or he may ask for clarification and then act on the new beliefs. In either case his participation in the class "discussion" may be quite different from the threatened closed-minded student.

To summarize this speculation I would like to list the main points of the argument:

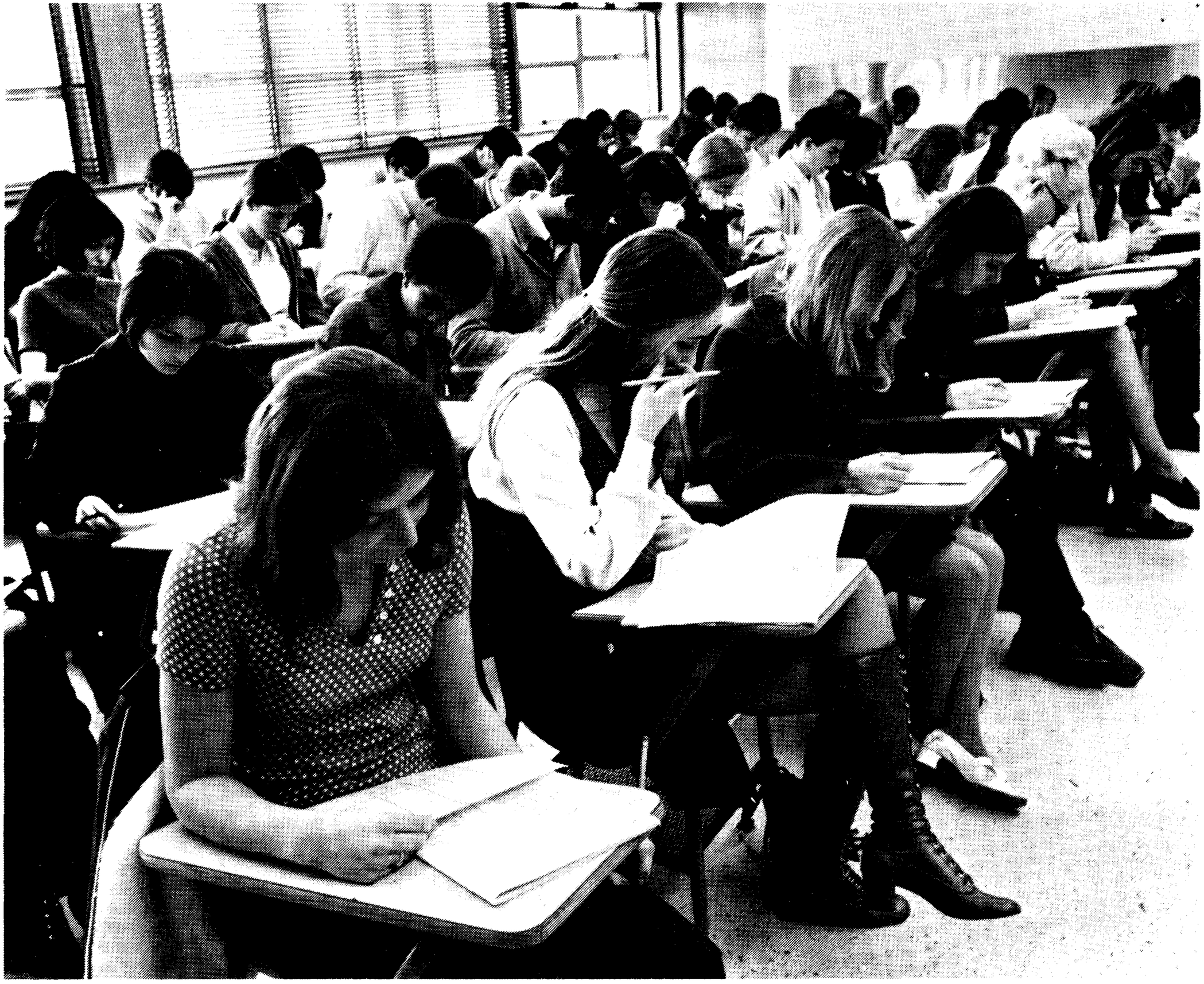
1. It is suggested that closed-minded students who are in low agreement with their instructor's goals act quite vocally in the classroom.
2. These vocal students become quite visible through their class participation and, as such, are looked upon with favor by their respective instructors.
3. The professor, regardless of whether he himself is open- or closed-minded, has a tendency to reward classroom participation by viewing vocal students as "above average," and then later by allowing such mid-semester estimates to influence the assignment of final grades.

Hopefully, the position offered in this paper will be clarified and strengthened or refuted through future research. However, in this interim period where we operate in the twilight zone of uncertainty, we might at the very least be aware of what we *may* be rewarding when we say, ". . . and finally, your grade will be influenced by your class participation."

References

1. Robert L. Ebel, *Measuring Educational Achievement.* New Jersey: Prentice-Hall, Inc., 1965.
2. Orville Palmer, "Seven Classic Ways of Grading Dishonestly," *English Journal,* October, 1962, pp. 464–67.
3. Milton Rokeach, *The Open and Closed Mind.* New York: Basic Books, Inc., 1960.
4. *Ibid,* p. 57.

5. *Ibid.*
6. Jacob W. Getzels and Philip W. Jackson, *Creativity and Intelligence.* New York: John Wiley and Sons, 1962.
   Eugene O. Alexander, "The Marking System and Poor Achievement," *The Teachers College Journal,* December, 1964, pp. 110–13.
7. Paul L. Dressel and Irvin J. Lehmann, "The Impact of Higher Education on Student Attitudes, Values, and Critical Thinking Abilities," *The Educational Record,* Summer, 1965, pp. 248–58.
8. Rokeach, *op cit.,* p. 107.
9. James A. Conway and John R. Dettre, "An Exploratory Study of the Relationship of Belief Systems, Goals, and the Evaluation of College Undergraduates," unpublished research sponsored by Research Foundation of the State University of New York, 1967.
10. Getzels and Jackson, *op. cit.*
11. Rokeach, *op. cit.*
12. James A. Conway, "Problem-Solving in Small Groups as a Function of 'Open' and 'Closed' Individual Belief Systems," *Organizational Behavior and Human Performance,* November, 1967, pp. 394–405.



Van Bucher

PART 6

The Evaluation of Learning

6.1 *The Role of Evaluative Research*

EDWARD A. SUCHMAN

Reprinted from *Toward a Theory of Achievement Measurement, Proceedings of the 1969 Invitational Conference on Testing Problems,* with permission of the publisher: Educational Testing Service, Princeton, New Jersey; Copyright © 1970. The late Edward A. Suchman was a member of the sociology faculty of the University of Pittsburgh and was active in educational and public health research.

The principles of evaluation are much the same, irrespective of whether we are evaluating educational processes, the effectiveness of large organizations, or a new program. In this paper, Suchman makes the point that measurement is not evaluation, but evaluational processes require that measurement take place. He relates his analysis of evaluation to a model much used by behavioral scientists these days—the idea that any organism can be perceived as an energy system that processes inputs and produces outputs.

The Current Demand for Evaluative Research

In times of rapid social change, traditional public service and social action programs are apt to find themselves under constant challenge from new and untested approaches. Which of the old should be revised or discarded completely and which of the new deserves a trial? How much change is necessary? Can the old be patched up and made to work or must wholly new policies and practices (16) be developed? The climate of such times is likely to be one of vigorous academic debate and public conflict concerning what steps need to be taken to meet the ever increasing cries of dissatisfaction and dismay (5). The presence of such dissatisfaction in almost all areas of public service today is perhaps the key to the current demand for evaluation. Evaluation feeds on dissatisfaction and change. Action, almost any kind of action, is frantically sought as a means of alleviating the discontent or, at least, of postponing open conflict. And, almost as an apology for too precipitous action, evaluation is often proposed as a means of maintaining rationality and control (7).

The field of education today, like that of health and social welfare, is under pressure to change its traditional programs and organization. Industrialization, urbanization, civil rights and minority movements, changes in educational technology, new occupational demands have created a strong public and professional search for new educational approaches. In many cases, trial and error, rather than carefully planned change, has characterized these educational innovations. For the most part, these new programs have been developed without any appreciable relevant theoretical basis (2). The less obvious the theoretical justification for a program, the greater will the need be felt to evaluate its success or failure. In fact, a major rationalization for a trial-and-error approach is that one can evaluate the trials and determine the errors. If the program can be made to work through "rank empiricism," the theoretical reason (if any!) for its success can be determined "after the fact" (1). Thus, too often, the situation becomes one of introducing innovation for innovation's sake alone, and perhaps it is exactly the highly trial-and-error nature of so many of these educational innovations that has led to the current demand for more intensive evaluation.

In short, the need for evaluation in education today is so great because we have lost faith in our traditional programs and are uncertain as to what we are trying to do with our new programs or why they should work at all. What the field of education (and, we might also add, health and welfare) lacks most today is an understanding of how to evaluate the relevance of its current fund of knowledge for social needs, and then utilize this knowledge in the development of new programs more suitable to these needs. This inadequacy stems, to a large extent, from the underdevelopment of what we might call "educational practice theory." Teaching as a profession has developed over the years without sufficient attention to the underlying rationale for

its many instructional practices. Based largely on untested assumptions, these practices have been handed down from one generation of students in teachers' colleges to another (8). While supposedly based on the psychological theories of learning, only recently have we become aware that a wholly different type of "instructional" theory is necessary for the practice of teaching as opposed to the psychology of learning (9).

The relevance of this distinction between basic theory and professional practice theory for evaluative research has recently been underscored by the report of the Special Commission on the Social Sciences to the National Science Board entitled *Knowledge Into Action: Improving the Nation's Use of the Social Sciences* (14). According to this report. "The professions are among the main social institutions through which social science knowledge can be translated into day-to-day practice" (15). Evaluative research, it is pointed out, provides one of the major avenues whereby the various fields of professional practice can clarify, develop, and test the underlying rationale for and theory of their professional practice. Program evaluation is the *sine qua non* of professional efforts to translate knowledge into action.

A Conceptual Approach to Evaluation

At the most general level, evaluation refers to the process of determining worth or value. This process may range from a subjective assessment based on personal experience to a highly controlled experiment based on the scientific method (11). The object of evaluation may be as broad as national policy or as narrow as some individual need. In most cases, however, the purpose of such evaluation is to direct future action. Thus, we can locate evaluation in the area of decision making whether such a decision involves some personal act or some major planned social change (13).

Given such a wide range of application, it is not difficult to understand why so much confusion and disagreement exists today as to what constitutes the "proper" method of evaluation. Different purposes will require different evaluation designs. At the moment we lack any agreed-upon systematic scheme for classifying or ordering different forms and types of evaluations. Furthermore, it is our opinion that the field of evaluative research is changing much too rapidly to make any high degree of formalization desirable at the present time. It is our hope, however, that the following remarks will at least indicate some of the major theoretical and methodological problems in this area.

First, it seems unwise to conceive of evaluation studies as requiring some special theoretical or methodological model of their own. The basic differences between evaluative research and nonevaluative research are the same as those between any form of applied research and basic research. These differences have received detailed discussion in a number of articles and need not be repeated here (10). Nevertheless, a great deal of the current controversy over both the "theory" and "method"

of evaluative research tends to lose sight of the fact that the major purpose of most evaluative research is administrative. The primary goal is usually to aid the decision-making process concerning some social problem or policy. Even in the conduct of evaluation studies, administrative considerations will often have precedence over scientific ones (12). Therefore, it does not seem productive to continue arguments over the degree to which the classical experimental design is applicable to evaluative research (3). As is so often the case in such controversies, the answer would be "It depends on the purpose" (4). In the early stages of evaluating a new program, it would probably be more profitable to utilize a rather fluid, clinical case study, "anthropological" type of design. Increased understanding of the problem and more detailed specification of the type of activity to be carried out could then be followed by a survey evaluation design which would provide preliminary evidence as to the effectiveness of one's program on an *ex post facto* or longitudinal basis. Finally, when the stage is reached for a definitive test of some particular program, it would then be possible to proceed to a more rigorous experimental design (18).

One cannot really argue in the abstract as to which approach is "correct." We have found it useful to distinguish between "pilot" projects where the main objective is to try out different approaches and where a flexible, anthropological approach provides the greatest amount of information and "model" projects where the emphasis is upon testing a program under ideal conditions and where a more rigorous experimental design is indicated. "Prototype" projects call for an operations research design whose main emphasis is upon the feedback of information for program improvement (19).

These kinds of practical considerations do not, of course, rule out the need to view evaluation research within a broader theoretical or methodological perspectivve. In our work, we have found it most helpful theoretically to link the evaluative research model to that of the independent-intervening-dependent variable sequence of multicausal analysis. In terms of this model, the program activity becomes the independent variable which is to be manipulated or changed. The intervening process represents the "causal" factors which promote or inhibit the development of the valued goal. The dependent variable then becomes the desired objectives or changed conditions. Given this formulation, the nonevaluative hypothesis "If A, then B" becomes the evaluative hypothesis "By changing A (through a planned intervention program), the 'causative' process B will be effected in such a way that the probability of producting effect C (which I judge to be desirable) will be increased." Underlying this hypothesis are several key theoretical issues: 1) What are the "causative" factors affecting the achievement of the desirable outcomes? What changes must one produce in the underlying process in order to bring about a change in the result? 2) What activities, programs, or techniques can one develop for deliberate intervention into this "causative" process so as to increase the prob-

ability of the desired outcome? What are the social, economic, political constraints limiting one's ability to utilize a particular intervention? and 3) What is the total picture of changes produced by the intervention? What negative (unintended) as well as positive (intended and unintended) effects take place? What other aspects of the social system (20), besides the one within which one is working, are affected?

In our opinion, these three sets of questions represent the key "theoretical" issues underlying evaluative research. They point very decisively to the need for the development of theoretical models which will link professional activities to desired social outcomes. Basically, this is what we mean by professional practice theory. As the report to the National Science Board has documented, this area has been thoroughly neglected in most professional school training programs. Specific activities as well as entire programs are carried out largely on the basis of tradition without sufficient attention to the rationale for believing that they are capable of producing the desired results. On the other hand, "academic" disciplines continue to develop basic theories which fail to tie into professional practice. Finally, national policy or goals are set by social and political forces without due consideration of the professional consequences as determined by professional groups. The model we propose inherently requires a close linkage between professional practice and academic theory.

This formulation of the evaluation problem also permits one to distinguish between what we might call a "technical" versus a "theoretical" failure. Professional service or social action programs may fail to achieve their desired goals either because they are operating according to an invalid underlying theory of process (a theoretical failure) or because, even though the rationale is correct, they cannot succeed technically in developing and implementing a program that successfully engages the underlying change process. Confusion of technical with theoretical failures underlies much of the current controversy regarding programs in almost all fields of health, education, and social welfare. For example, does the apparent failure of programs such as Head Start stem from an invalid theory concerning the effect of early environmental deprivation upon intelligence or can it be more correctly explained in terms of a failure to institute the proper type of early intervention programs?

The preceding formulation of the evaluation problem, incidentally, also indicates the crucial contribution which evaluative research can make to underlying theoretical understanding. Professional practice and social action can serve as a highly significant crucible for the testing of academic theory.

Levels of Evaluation

Evaluative research may further be viewed as taking place on a number of different levels, each requiring a somewhat different conceptualization of the evaluation problem. We may order these levels from the broadest to the most narrow as dealing with (a) social systems; (b) organizations or institu-

tions; (c) programs or projects. The major emphasis of the social-system level of evaluation will be upon policy or ultimate objectives, while an organizational evaluation is more likely to be concerned with intermediate administrative objectives and a program evaluation with immediate service objectives. On the social-system level, the objectives most likely would deal with ideological or value questions and be aimed at policy decisions of a social-political nature. Evaluation at this level challenges the goals and assumptions of the major societal subsystems in terms of their ability to affect such fundamental social values as the public's health, education, and welfare. Evaluative research strategy at this level tends to be more descriptive and philosophical than empirical and scientific, involving "great debates" over national goals and the means towards these goals. Objectives and criteria are usually formulated in terms of gross social indicators of "progress" towards overcoming disease, ignorance, and poverty. Social "bookkeeping" provides the basic evaluative data on effort and accomplishment, with political judgments being made concerning the adequacy and efficiency of comparative systems. Examples of evaluation studies at this broad level would include such comparisons as the "socialized" versus "free enterprise" medical care system, the educational approaches, a welfare system based on need or entitlement.

At the next lower level, arganizational evaluations have as their objective the evaluation of the structure and operation of the major institutional arrangements whereby the broad social goals of a society are to be pursued. On this intermediate level of administrative evaluation, we are most concerned with a systems approach for evaluating the overall operation of major organizations such as the Department of Health, Education, and Welfare. This type of evaluation would stress the objective of providing "accountability" information for organizational management. This level of evaluation is becoming increasingly important as massive social programs are mounted involving major federal, state, and local public service agencies. Evaluation at this level has as its major objective staff policy formation and change involving the allocation of resources and the assignment of priorities; it is apt to have the greatest impact upon planned social change. Long-range objectives are more likely to be affected by organizational or system changes than by specific program innovations. In fact, one of the major criticisms of much current evaluation aimed at the program level is that this form of "tinkering" is too specific and symptom-oriented to really ameliorate any major social problems.

Finally, on the level of immediate objectives, we have what is most common in evaluative research—program evaluation. Evaluation studies at this level are apt to concentrate on the "effort" category. Specific programs become evaluated largely in terms of input, or the amount of effort expended rather than the actual accomplishments of the programs. The objectives at this level stress criteria as the quality and quantity of personnel, performance ratings, and amount of service rendered. It is usually taken for granted that such ac-

tivities, if successfully mounted. will *ipso facto* produce the desired outcomes. Such operational programs are usually fairly easily evaluated by means of an experimental or quasi-experimental design. In fact, this is probably the most appropriate level of evaluation for the use of such designs. It is extremely important, however, to remember that an evaluation at this level should not formulate objectives which involve more than a measure of the accomplishment of the program being evaluated. Such single programs are not likely to show any appreciable impact upon intermediate or ultimate objectives. The inclusion of such higher-level objectives in the evaluation of service programs is probably the main reason why so many evaluations appear to indicate failure. It is a very rare service program indeed that can show any impact beyond the immediate level of its specific operational goal.

Evaluation and Measurement

Since the major focus of this conference is upon measurement, perhaps a brief word is in order concerning the relationship of measurement to evaluation. Our approach to evaluation stresses the testing of some hypothesis concerning the relationship between planned activities and desired objectives. This relationship has been referred to in terms of a broader input–process–output model. Stated in this form, it becomes apparent that measurement is an inherent aspect of evaluation research. Most critically, perhaps, such measurement refers to the conceptualization, isolation, and measurement of specific criteria representative of the relative degree of success or failure in the attainment of the desired output. In general, this has been referred to as the problem of criterion measurement and has a long and interesting history. Some would maintain that until one has developed reliable and valid criterion measures for the desired objectives, evaluation cannot take place in any rigorous, scientific way. The time-honored question "Validity for what?" presents a critical problem for evaluation research. All programs will have some effect and unless such efforts are specified in advance and clearly related to the desired values or goals, evaluation research stands in danger of simply showing that some kinds of effects took place. Similarly, it would seem important that the measurement of effect embrace the magnitude and duration of the effect both in terms of the costs involved and the degree of effect necessary to make a program worth undertaking.

The problem of criterion measurement, however, is much more complicated than would appear from a review of the technical aspects of criterion validity. As formulated in this paper, the major objective of an evaluation study is to aid in the decision-making process. The rightness or wrongness of such decisions are only slightly amenable to evaluation by current educational tests and measurement (6). How, for example, does one measure the effectiveness of a social policy concerning welfare? The development of so-called social indicators which reflect the well-being of a population go far beyond the standard techniques of achievement

tests and measures (21). Social goals and objectives are not easily translatable into specific criterion measures. It is much more likely that such evaluations will relate more to progress toward an ever-changing goal than to achievement of a specific goal.

This point is also relevant to the other major components of our input-output model. It seems to us that for many types of evaluation, a much more meaningful and realistic form of measurement would concern itself with the developmental process itself rather than with the final output. Thus, we face a wholly new set of measurement problems in terms of process. An important element in the definition of such process criteria is an understanding of the underlying sequence of events leading toward the desired ultimate objective. Given such a sequence, it is possible to then determine certain crucial points along a time line which could be used as evaluative measures of progress (17).

Finally, we come to the measurement of the "input" component of the evaluation process. Measurement in this case becomes essentially one of defining, and varying, the essential program components that constitute the activities of the intervention or action program. The crucial question is "How does one measure the degree to which one has successfully introduced the type of program that is being hypothesized as the desired treatment? As we have stated previously, unless one has a reliable and valid measure of input, it is impossible to determine whether any future failure was due to the inadequacy of the underlying theory or to the technical inability to put into effect the desired program. In its most desirable form, it would be possible to vary the input and to relate such variation to differing degrees of process involvement and output production. Thus, the problem of measurement of program components becomes an extremely important one for evaluative research.

In summary, measurement may or may not involve evaluation; however, all evaluation implies some form of "measurement" of all three components of the input-process-output model.

References

1. Bloom, Benjamin. Some theoretical issues relating to educational evaluation. In Ralph W. Tyler (Ed.), *Educational Evaluation: New Roles, New Means.* Chicago: University of Chicago Press, 1969. pp. 26–50.
2. Brickell, Henry M. Appraising the effects of innovations in local schools. In Ralph W. Tyler (Ed.), *Educational Evaluation: New Roles, New Means.* Chicago: University of Chicago Press, 1969. pp. 284–304.
3. Campbell, Donald T. and Stanley, Julian C. Experimental and quasi-experimental designs for research on teaching. In N. C. Gage (Ed.), *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963. pp. 171–246.
4. Chapin, F. Stuart. *Experimental Designs in Sociological Research.* New York: Harper and Bros. 1947 (revised 1955).
5. Conant, James B. *Shaping Educational Policy.* New York: McGraw-Hill, 1964.
6. Flanagan, John C. The uses of educational evaluation in the development of programs, courses, instructional materials and equipment, instructional and leaving procedures, and administrative arrangements. In Ralph W. Tyler (Ed.), *Educational Evaluation: New Roles, New Means.*

Chicago: University of Chicago Press, 1969. pp. 221–241.

7. Freeman, Howard and Sherwood, Clarence C. Research in large-scale intervention programs. *Journal of Social Issues,* January, 1965, 11–28.
8. Glaser. Robert. Ten untenable assumptions of college instruction. *The Educational Record,* spring, 1968. 154–159.
9. Glaser, Robert. Theory of evaluation of instruction: changes and trends. From the *Proceedings of the Symposium on Problems in the Evaluation of Instruction.* Los Angeles: University of California, December, 1967.
10. Hemphill, John K. The relationship between research and evaluation studies. In Ralph W. Tyler (Ed.), *Educational Evaluation: New Roles, New Means.* Chicago: University of Chicago Press, 1969. pp. 189–220.
11. Herzog, Elizabeth. *Some Guidelines for Evaluative Research.* Washington, D. C.: U. S. Department of Health, Education, and Welfare, Social Security Administration, Children's Bureau, 1959.
12. Hesseling, P. *Strategy of Evaluation Research.* Netherlands: Van Gorcum and Co., 1966.
13. Hyman, Herbert H., Wright, Charles R. and Hopkins, Terence K. *Applications of Methods of Evaluation: Four Studies of the Encampment for Citizenship.* Berkeley: University of California Press, 1962.
14. *Knowledge into Action: Improving the Nation's Use of the Social Sciences.* Washington, D. C.: National Science Foundation, 1969.
15. *Ibid.* p. 21
16. Miles, Matthew B. *Innovation in Education.* New York: Bureau of Publications, Teachers College, Columbia University, 1964.
17. Provus, Malcolm. Evaluation of ongoing programs in the public school system. In Ralph W. Tyler (Ed.), *Educational Evaluation: New Roles, New Means.* University of Chicago Press, 1969. pp. 242–283.
18. Suchman, Edward A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs.* New York: Russell Sage Foundation, 1967.
19. Suchman, Edward A. Action for what? A critique of evaluative research. Paper prepared for symposium on *The Organization, Management, and Tactics of Research.* Vocational Guidance and Rehabilitation Services, February 20–21, 1969. (In press).
20. Suchman, Edward A. A model for research and evaluation on rehabilitation. In Marvin Sussman (Ed.), *Sociology and Rehabilitation.* Washington: Vocational Rehabilitation Administration, 1966. pp. 52–70.
21. Sheldon, Eleanor Bernert and Moore, Wilbert E. *Indicators of Social Change.* New York: Russell Sage Foundation, 1968.

6.2 *The Functions and Uses of Educational Measurement*

WINTON H. MANNING

Reprinted from *Toward a Theory of Achievement Measurement. Proceedings of the 1969 Invitational Conference on Testing Problems,* with permission of the publisher: Educational Testing Service, Princeton, New Jersey; Copyright © 1970. Winton H. Manning is the director of the Division of Developmental Research of the Educational Testing Service.

Manning reviews the uses that educational measurement has served in the past and points to two other functions of education in which it is playing an increasingly important role: the development of theories, particularly those related to problem solving, and the individualization of learning through self-discovery. His concept of measurement is not the traditionally narrow and confining one, but rather one that expands and diversifies — much as education itself is expanding and diversifying these days.

The main argument presented in this paper is that functions and uses of educational measurement that have developed in the past are insufficient for the future because they have been too much shaped by the practical problems of educational institutions to the neglect of other functions. The institutional problems that have demanded priority in the past are primarily those that arise from our having viewed the educational system as a training resource designed to supply the manpower needs of industry. Although this aspect cannot be wholly overlooked if we are to maintain a highly technological society, it is nevertheless unsatisfying, particularly if we believe that educational measurement offers the best means we have for effecting improvement in the quality of the educational process. Aside from the rhetoric that is employed, differences in the uses of tests in education as compared with industry or the military are probably more apparent than real, and this fact alone should signal that something is wrong.

If there is merit in this argument, then it follows that we should be particularly concerned with identifying and fostering the development of measurement functions in education that are uniquely appropriate to the needs of young people and to the traditions of rational, objective, scientific inquiry into the process by which young people are educated. Put another way, two issues that arise when one examines the functions of educational measurement are:

1. Because educational measurement is oriented mainly toward the solution of practical problems of educational institutions, its functions have been those of providing a means for accomplishing certain tasks of social and educational engineering—that is, successively sorting people into hierarchies of talent and accomplishment for the world of work—rather than as an instrument in the construction of educational theories that are amenable to scientific investigation.

2. Educational measurement has been mainly employed in the solution of problems confronting educational institutions as they seek to shape human resources for economic development, rather than in the solution of problems of individuals as they seek to use the resources of educational institutions for self-development. As a consequence, testing has not progressed as far as it should as a means for assisting students to encounter successfully those problems of self-understanding, choice, and decision making that they confront as maturing individuals in a modern technological society.

This way of looking at testing suggests there exists an imbalance in the development of educational measurement—a practical, or a-theoretical, bias on the one hand, and an institutional, or a-personal, bias on the other. In taking this view, my intention is not to disparage the traditional uses of educational measurement, or even to view these as becoming less important in an absolute sense in the future. However, forces do exist that seem likely to extend the functions of tests into new directions in the future. Two of these are:

First, opportunities afforded by educational technology—particularly computer-assisted instruction, testing, and guidance—seem likely to make the development of scientifific educational theory more feasible and, hence, more attractive. Correspondingly, if we are wise enough to use it that way, the new cybernetic technologies also offer us the opportunity to devise measurement procedures that more effectively serve the individual student for purposes of his own development. In brief, we have potentially within our grasp the means for employing measurement directly in the interests of students, rather than indirectly through the presumably beneficent ministrations of mediating institutions or agencies.

Second, the revolution in values, attitudes, and beliefs of young people throughout the world augurs for profound changes in educational institutions at all levels, and as a consequence, the way in which educational measurement is employed may shift from the traditional institutional concerns to new problems and, hence, to new functions. I believe these revolutionary forces may well serve to undermine the practical and institutional biases that presently too much characterize the functions and uses of educational measurement; or to be more exact, I hope that this will be one of the consequences.

With this general introduction, let me turn to a brief discussion of some of the older uses of measurement in education and to a consideration of some functions that seem likely to grow in importance.

Traditional Uses of Educational Measurement

As Henry Dyer (4) pointed out several years ago, we can readily identify at least three important functions that educational tests have been designed to serve in the past. These are:

1. *Selection and distribution* in which tests are used as a basis for the selection of students for programs, or the distribution of students among programs, and where the distribution system is aimed at providing an optimal match between student abilities and limited educational resources
2. *Diagnosis or prescription* where tests are used as a basis for indentifying the nature and extent of educational deficiencies, and for prescribing educational treatments designed to remedy these deficiencies, thus aiming at maximizing the number of children reaching a given level of achievement
3. *Evaluation* where tests are used to assess the effectiveness of educational programs so that there is a systematic basis for comparison of educational outcomes and, hence, for the improvement of educational practices

It would be tempting to consider how the uses of tests for selection, diagnosis, and evaluation have evolved over the past 50 years and to speculate on how these functions are likely to change in the decades ahead. At the level of higher education, for example, demands for open admissions to col-

lege are causing not only an intensive study of the effectiveness and equity of tests that are used in selection but also a profound re-examination of the moral and ethical bases for the process of selection itself (12). Similarly, the concern for equalizing educational opportunity has led to enhanced demands for tests that diagnose educational deficits, particularly of disadvantaged, minority children, in ways that will permit prescriptive rather than random efforts at remediation (5, 10). Finally, the concern with evaluation of educational programs has grown dramatically in the past decade, partly as a consequence of the promise that this approach holds for improving the educational process and partly as a result of the increased demand for accountability in education (11).

There is little question that these functions of tests will continue to be among the major pivots around which measurement research and testing programs will continue to revolve in the years ahead. However, there are other uses of educational measurement that are important and which I should like to call to your attention: 1) the uses of tests in the development of educational theory, and 2) the educative or guidance function of educational measurement.

Educational Tests as Instruments of Educational Theory

Measures of educational outcomes may be conceptualized as real traits that are acquired as a consequence of instruction. Aside from determining whether the sample of behavior displays reasonable consistency, the validity of such measures is established by assessing the extent to which the proposed interpretation of the test corresponds to some real trait or, in other words, the investigation of the construct validity of the measure. To do this requires study of the content of the items, the interrelationships or structure that exists among the items, and the relationships of the test responses to behavior that is manifested external to the test itself. In Loevinger's analysis (9), for example, the three components of construct validation are described in terms of: 1) the substantive validity of the test (which for achievement tests may be seen as equivalent to the problem of content validity); 2) the investigation of the structural validity of the tests or the extent to which test items parallel the structural relations of other manifestations of the trait being measured; and 3) the external validity of educational tests, or the degree to which the test is related to behavior displayed outside the testing situation.

Seen from this perspective, the development of a theory of educational achievement is not materially different from the task of developing personality theory or psychological theory. The problem of educational testing as a theoretical instrument rather than as an applied technique is, it seems to me, central to many of our other concerns, and is indeed the basis for the motivation of most behavioral scientists who choose to study education.

The concept of a "trait" is, of course, essential to the whole operation, and

it is here that educational theory confronts some particularly thorny problems. The objectives of education, when stated in terms of specific behaviors, often focus more upon products than upon processes of behavior. Although the taxonomies of Bloom (2) and Krathwohl (7) appear to be much concerned with assessing the various ways of acquiring and using knowledge, there is reason to assert that educational measurement practices tend to subordinate the question of describing *how* one goes about seeking a solution to a problem, and to enhance the goals of teaching students to display acceptable solutions. Whether this is so or not, there is little question that educational measurement has been mainly developed as a technique for evaluating the quality of outcomes, rather than for describing the characteristic strategies that individuals use in reaching those outcomes.

The reasons for this are many, including particularly the origins of educational testing practices within a meritocratic selection framework, and the noticeable tendency in educational research to avoid studying behavior that depends as much upon affective as upon cognitive processes. Furthermore, if progress through the educational system is seen as a competitive race for the rewards that society can bestow on the successful, there is reason enough to understand why we have to emphasize measurement of traits of ability or accomplishment to the exclusion of other interesting characteristics of the learner.

At the level of early childhood education, the contrast is quite evident when one compares the behavior that is sampled by conventional intelligence tests and the approach to measurement that, stemming from Piaget's work, has characterized the project known as *Let's Look at Children* (8). Here the emphasis is not on whether the child has learned to perform such acts as stringing beads or defining words, but rather upon understanding the kinds of cognitive processes that the child uses when he is confronted with an interesting and challenging problem that requires an explanation of the way in which he sees his world.

Obviously, a major deterrent to the development of tests that permit us to observe problem-solving strategies has been the awkwardness and expense of such procedures. For example, tab tests and similar approaches have not seemed, in most cases, to be dramatically better than conventional tests for conventional purposes, and, hence, it is difficult to justify their far greater expense.

The great potentiality of the computer as a medium for testing seems to be in the capacity it affords for real-time interaction with the subject and the consequent ability we acquire to record objectively the strategy that the individual uses in seeking a solution as well as its outcome. However, as some have pointed out, serious questions arise when one considers computer-based, problem-solving tests. What aspects of behavior we should look at, and how these observations should be combined to yield quantitative or qualitative descriptions that are useful, are not always immediately evident. For most persons, the characteris-

tic way in which we go about searching for solutions to a problem may have more significance for the later performance of social and occupational roles than the degree to which we have mastered the content of a discipline or subject in school. From the standpoint of education, the development of tests that are oriented toward assessment of problem-solving styles would have, therefore, in my judgment, a salutory effect on educational practices. Furthermore, the possibility is growing that educational objectives will again come to be specified in terms of narrow product-oriented behaviors, leading ultimately to a kind of neo-positivism of the classroom. An important deterrent to such excesses of naive behaviorism may be that, through development of measures of both process and outcome, we will be able to construct theories of human behavior that will lead to more faithful statements of educational objectives and, hence, to more fruitful hypotheses about educational treatments. Richard Atkinson (1) has suggested that the lack of a real theoretical foundation for the teaching of reading was not clearly evident until the problem of devising a program for teaching reading to children by means of computer-assisted instruction was confronted. I hope that something analogous to this may occur to theories of intellectual growth when we confront the problem of assessing problem solving within the framework of a computer-based test.

The search for consistencies in problem-solving strategies and the attempt to devise and test educational theory through construct validation of such tests will require as much attention to Loevinger's second and third components—structural validity and external validity—as to her first component, substantive or content validity. Among other benefits, not the least would be that of placing educational measurement squarely within the same methodological domain as that which has characterized psychological test theory for some years.

In dwelling briefly on the function of educational tests in theory construction, I have called your attention to a function of tests that have always been recognized but has been more often a pretension than an actuality. My reason for emphasizing this use of tests derives primarily from the widely shared conviction that in the coming decade educational programs are likely to change in dramatic ways not only as a consequence of technological developments associated with the computer but also in response to revolutionary social forces aimed at restructuring social institutions. Greater attention to the problem of educational theory and the relationship of educational tests to theory would, therefore, seem to be a productive and stabilizing influence in the turbulent years that lie ahead.

Educative and Guidance Functions of Measurement

Some future social historian may well characterize this century as one of great optimism, in which the increased perfectability of the human condition was assumed without question, and where science and technology were seen by civilized men as the means for

bringing about continuous human advancement. Education has, if anything, embraced this view with even greater ardor than other social institutions, with the ultimate consequence that something like one quarter of a billion tests are administered each year to students enrolled in our educational institutions. In many respects this enthusiasm for tests seems to rest more upon blind faith than upon observable benefits to the consumer. In nearly every school, for example, the yellowing pages of unused but carefully stored printouts of test score rosters bear mute witness to a problem that is characterized less as an information overload than as the totemism of test use.

However, the educational uses of computer-based information storage, handling, and retrieval systems coupled with the prospective establishment of education networks offer the possibility of extending the concept of a test to include more effective interpretations of the meaning of the test performance. For many years the need to provide better interpretative information than norms, grade equivalents, expectancies, or probabilities of success has been recognized, but despite the good intentions of testing agencies, the rate of misuse or disuse of test results has continued to mount.

The concept of educational measurement embedded in a system that incorporates a delivered interpretation of the meaning of the test performance will require some profound adjustments in testing and guidance practices. As such systems develop, questions arise as to what boundaries, if any, exist between the uses of tests in counseling and guidance and the technology of educational measurement itself. Conceiving of tests as components in computer-based measurement and guidance systems leads to the assumption that the boundaries must become blurred, if not nonexistent. It is, therefore, exactly this kind of fusion of functions that I am pointing toward in discussing still another use of educational measurement—namely, *the educative or guidance uses of tests.*

The educative use of tests suggests that evaluation of student achievement, attitudes, and values through educational measurement should carry with it the obligation to portray the individual in terms that will permit him to learn more about himself through an analysis of his own performance rather than primarily through comparisons with other groups of students. The uses of tests for institutional ends, such as selection, have fostered a notion that evaluation of performance is only possible under conditions of competition, or what B. A. Thresher called "adversarial" testing. But if the use of test information is centered on the processes of self-understanding and self-discovery by the student, achievement testing that is epistemological rather than adversarial would seem to be the natural result. A good example of this is provided in computer-assisted instructions where we see the development of criterion-referenced tests that are embedded in the interactive instructional programs.

Applying this viewpoint to systems of guidance information, such as those being developed by Martin Katz at Edu-

cational Testing Service and David Tiedeman at Harvard, raises issues concerning the function of most forms of educational testing in the future. For example, consider the familiar question of how test information can be interpreted "so that students can choose among various alternatives more realistically." As Katz points out, the question itself may be a false or misleading one, for a person rarely confronts a situation in which there is truly a fixed set of alternatives; rather "he often has some opportunity to construct or create his own options" (6). Furthermore, there is, as Katz eloquently describes, an enormous difference between seeing the task of guidance as that of helping students make "wise" or "realistic" decisions through choosing the best alternatives—in other words, those with highest probability of pay-off—and the task of assisting students to become wise in the processes of decision making. In this sense, there is an interesting correspondence between emphasis on *processes* in educational theory and a comparable emphasis in guidance theory.

Whether we are talking about computer-based systems of guidance that are student-centered, or about measures of problem-solving strategies as instruments of educational theory, the observations made by Jerome Bruner concerning the kind of education we need for the future may be relevant. Writing a year ago, Bruner (3) said:

"...we shall probably want to train individuals, not for the performance of routine activities that can be done with great skill and precision by devices, but rather to train their individual talents for research and development which is one of the kinds of activities for which you cannot easily program computers. Here, I mean research and development in the sense of problem solving... What this entails for education is necessarily somewhat obscure, but its outlines may be plain. For one thing it places emphasis on the teaching of interesting puzzle forms—ways of thinking that are particularly useful for converting troubles into problems ...for converting chaotic messes into manageable problems..."

If tests are to become integrally embedded in information systems designed to foster a sense of planfulness, orderliness, and continuity within the lives of students—or, in other words, if measurement is to assist in the process by which civilized men seek to convert into manageable problems the chaotic messes that ignorance of self and powerlessness seem to decree for them—the mission for educational measurement is indeed formidable. Whether educational measurement can become as proficient a servant of individual human beings as it has been the handmaiden of educational institutions is an important issue that deserves serious debate and creative energy.

Conclusion

In summary, we have identified three major functions of educational measurement that have developed in the past: selection, diagnosis, and evaluation. It was asserted that these uses of

tests arose primarily from institutional needs of the educational system although their use by institutions may indirectly have also served the needs of students. Two functions of tests that deserve particular emphasis at this time are: first, the uses of educational tests in the construction and evaluation of educational theories, especially theories that give particular attention to processes or strategies of problem solving rather than outcomes alone; and second, the uses of tests in the service of individual students through systems of guidance that employ measurement as a means of fostering self-discovery and as a means for encouraging students to develop wisdom in decision making.

Development of these testing functions will require that educational measurement become integrally involved in both instruction and guidance, particularly in those approaches that utilize the unique capacities of the computer for interaction and objectivity. The search for means of expanding the functions of testing within the context of educational technology may also have the effect of reinforcing a humane use of modern technology rather than simply extending the mechanical efficiency of present functions of educational measurement.

Four hundred years ago Montaigne wrote of education in his day:

We labor only to stuff the memory, and leave the conscience and understanding unfurnished and void. Like birds who fly abroad and forage for grain, and bring it home in beak without tasting it themselves to feed their young, so our pedants go picking knowledge out here and there . . . holding it out at tongue's end, only to spit it out and distribute it abroad.

Pedantry is not confined to the classroom; it can exist within the confines of a machine-scorable answer sheet as well. New uses for tests in the years ahead must not simply be passive responses to the needs of the educational system, but energetic efforts to extend and elevate the functions of educational measurement.

References

1. Atkinson, Richard C. Computerized instruction and the learning process. *American Psychologist,* 1968. **4**, 225–239.
2. Bloom, B. S. (Ed.) *Taxonomy of Educational Objectives.* Handbook I: Cognitive Domain. New York: Longmans, Green and Co., 1954.
3. Bruner, J. S. Culture, politics, and pedagogy. *Saturday Review,* May 18, 1968. 69–72, 89–90.
4. Dyer, H. S. The functions of testing—old and new. *Testing Responsibilities and Opportunities of State Education Agencies.* Proceedings National Testing Conference for State Education Agency Personnel. The University of the State of New York, The State Education Department, Division of Educational Testing, Albany, New York. 1966. 63–79.
5. Gordon, E. W. and Wilkerson, D. A. *Compensatory Education for the Disadvantaged.* New York: College Entrance Examination Board, 1966.
6. Katz, M. R. Can computers make guidance decisions for students? *College Board Review,* No. 72, Summer 1969. 13–17.
7. Krathwohl, D. R., Bloom, B. S., and Ma-

sia, B. B. *Taxonomy of Educational Objectives.* Handbook II: Affective Domain. New York: McKay, Inc., 1964.

8. *Let's Look at Children.* Princeton, N. J.: Educational Testing Service, 1965.
9. Loevinger, Jane. Objective tests as instruments of psychological theory. *Psychological Reports,* 1957, **3**, 635–694. (Monograph Supplement No. 9).
10. Manning, W. H. The measurement of intellectual capacity and performance. *The Journal of Negro Education,* Summer, 1968.
11. Messick, S. *Evaluation of Educational Programs as Research on Educational Process.* Proceedings of the American Psychological Association, September, 1969.
12. Wilson, L. *Merit and Equality in higher Education.* Presented at the Annual Meeting of the American Council on Education, Washington, D. C., 1969.

6.3 *Testing Programs —Misconception, Misuse, Overuse*

FRANK B. WOMER

Reprinted from the *Michigan Journal of Secondary Education,* 1962, **2,** 153–161, by permission of the author and publisher. Frank B. Womer is staff director of the National Assessment Project and is a member of the educational psychology faculty of the University of Michigan.

Womer's article is concerned, as the title indicates, with the misuse and overuse of tests and misunderstandings about test results. Probably no instructional device is more used by teachers and least understood by them than is testing. Even the use of standardized tests is no guarantee against misuse and misunderstanding, as Womer shows.

We are in a boom period of standardized testing in elementary and secondary schools. Millions of tests are administered each year to pupils at all grade levels—achievement tests, mental ability tests, aptitude tests, and interest inventories, as well as several other types of tests and inventories. Some of these tests are given for college scholarship purposes and some for college admissions purposes. Title V of the National Defense Education Act has stimulated, and in some instances required, additions to testing programs at the secondary level. In general, however, these external influences account for a relatively small percentage of the total standardized testing undertaken by a school system.

There are at least two factors which have had a greater impact upon the amount of testing done in the schools than NDEA or college requirements. First, there has been and continues to be a natural growth of standardized testing at all grade levels. Second, the rapid growth of the guidance movement has meant a corresponding rapid growth in testing. This latter influence may well be the most influential one operating, for in many schools the testing program is developed by and operated by guidance personnel.

Inauguration or expansion of a testing program is relatively easy. Decisions made one day can be implemented within a week or two. The only time lag is that of the United States mail in delivering orders for tests and getting test materials from the publisher to the school. Machine methods for test scoring have reduced, and in many cases eliminated, objections that a testing program is a burden upon individual teachers. The school budget and allocation of time for testing are the only real problems to face if an administrator or faculty decides to enlarge the testing program. Thus it is relatively easy to test.

The values of standardized testing, however, cannot be dismissed so quickly. Such values are dependent upon two processes: (1) establishment of proper goals of testing and the development of a testing program to meet those goals, and (2) proper use of test results. Both of these processes are essential to the operation of a successful testing program. Most educators feel, and rightfully so, that the major weakness of testing today is in the area of test use.

Many writers have made this point, and most school administrators are acutely aware of the fact that the ultimate criterion for judging the effectiveness of their testing programs is the correct use of test results. Accumulating test scores in cumulative files is not evidence of test use. Correct use depends upon getting test results into the hands of counselors, teachers, administrators, pupils, and parents *and* of being sure that each consumer of these results is made knowledgeable enough to interpret them. In this latter statement—"made knowledgeable enough to interpret them"—lies the key to proper test use.

There are a number of ways that test scores are misused or overused and a number of misconceptions about tests and test scores that are common enough to warrant special attention. While one could think of innumerable

examples of specific errors in test interpretation, the purpose of this article is to point up some of the more common mistakes, in order to help increase the knowledgeable use of test results. Ten points have been selected for discussion; others could have been added.

Categorizing a Pupil at a Specific Level of Achievement or Ability

One of the most common mistakes made by persons unskilled in interpreting test results is the assumption of perfect reliability of a test score. Too often it is assumed that an IQ of 105 represents performance definitely superior to that represented by an IQ of 104 and definitely inferior to that represented by 106. Too often we fail to realize that a test score is best interpreted as a good estimate of the *general level of performance,* and that it will vary from test to test and from time to time. Test users must accept the concept of variability of test scores over time and over tests. The assessment of human traits and abilities is not at the same level of accuracy as that found in a physics laboratory. It probably is closer to the level of accuracy found in the predictions of weather, in which temperature predictions are generally within a few degrees of actual temperatures, but in which differences of ten or more degrees are common enough to be remembered vividly by critics.

Another aspect of this assumption of greater accuracy than actually prevails is the use of a single estimate (test score) to predict human performance. It is generally wise to insist on having two or three reading scores, or two or three aptitude scores, before putting much confidence in them. This is a direct result of the unreliability present in all test scores. If a pupil receives percentiles of 35 and of 40 on arithmetic tests given in two different years, one can have greater confidence that his level of achievement in arithmetic is in the average range than if only one of these scores is available.

Confusion of Norms and Standards

Norms are test scores which tell us the level of performance attained by an average or typical group of pupils. Standards represent human judgments of the level of performance that "should be" attained by a group of pupils. A test user should *not* assume that "typical" pupil performance is automatically the "proper" level of performance for pupils in a particular school system. It is reasonable, of course, to assume that pupils in many school systems will tend to perform at a level close to the level of test norms. In others, however, it is reasonable to assume that pupils will perform at a higher level or at a lower level.

One occasionally finds a test user who completely fails to grasp the conception of what a test norm is. Since a test norm represents "typical" performance, then, of necessity, *half* of the pupils in a typical or average group will have scores at or below the average score. If a teacher of a typical group of pupils finds that 40 per cent of his pupils are below grade level in reading, he is to be congratulated. In the norm group for whatever test is

being used, 50 per cent of all the pupils were at or below grade level. The assumption that all pupils in a class should be *at* grade level is patently impossible, unless one knows that the poorest achieving pupils in one's class is in the top 50 per cent of all pupils his own age or grade.

Assumption That Test Scores Predict Success or Failure for Individual Pupils

One way that test results often are overused is the assumption that a particular score or series of scores does in fact predict success or failure with unfailing accuracy. It is well established that students who succeed in colleges of engineering generally make high scores on numerical ability tests. Yet it is not correct to conclude from such data that Johnny, with a 50th percentile rank on a test of numerical ability, will not succeed in an engineering course. It is correct to conclude that of every one hundred students with numerical ability scores the same as Johnny's only a small percentage will succeed in an engineering curriculum. The test score does provide information of a probability type; it enables a student or parent or counselor to know the odds for success or failure. It is a well-known fact that long shots occasionally win the Kentucky Derby, but year in, year out, the favorites generally win.

It is not unusual for two counselors to look at the same test scores for an individual pupil and to come to somewhat different conclusions. For this reason, it is well to face the fact that while test scores do provide information that can be helpful in decision making, the decisions for courses of action are made by human beings, not by the scores.

Added to this overuse of test scores is the failure of some people to utilize all pertinent data available about a student when test scores are known. To allow test scores to outweigh all other judgmental data is a misuse of these scores; to ignore test scores in favor of other judgmental data also is a misuse of these scores.

Determination of Vocational Goals

"Mary's scores from a clerical speed and accuracy test and from a spelling test are only average. Therefore, Mary should not consider secretarial work as a career possibility." Or, "Since Jim's interest profile shows high scores in 'Scientific' and 'Social Service' he should elect a premed course in college." How often can vocational counseling be summed up in just such simple statements? It is so easy to make the jump from test score to occupation, and it seems so logical that this type of interpretation should be accurate. Unfortunately, the predictive validity of test scores in high school for success in specific occupations is not good enough to permit such interpretations. Most evidence of the predictive efficiency of test scores relates those scores to academic curricula. We can say with a fair degree of accuracy that certain patterns of test scores predict fairly well in different curricula. That is the type of validity data that is generally available.

The use of test scores in vocational counseling should tend to open doors of possible occupations rather than close them. Again, presenting the relationship between test scores and occupational areas on a probability basis can be helpful, and is certainly more accurate than making the assumption that certain test scores assure success in one field and failure in another.

Assumption That Intelligence and Achievement Are Separate and Distinct

Here are two sample questions from standardized tests:

1. Extraneous
   a. extra b. foreign
   c. transparent d. noisy
2. Make *indelible* means
   a. indistinct b. permanent
   c. purple d. identical

Both are vocabulary items. One of them is taken from a widely used intelligence test (California Test of Mental Maturity) and the other from a widely used achievement battery (Iowa Tests of Educational Development). Vocabulary items measure the learned meanings of words; vocabulary items are our best single measure of general intelligence or scholastic aptitude. Arithmetic items and general information items are also found in both achievement tests and intelligence tests. It is true, of course, that some items suitable for an intelligence test (number series, verbal analogies) are not good measures of achievement. It also is true that many direct measures of achievement (capitalization, punctuation, spelling) are not good measures of intellectual potential.

There is considerable overlap between standardized tests of achievement and standardized intelligence tests. One of the important differences between the two is the way the results are used. When analyzing achievement test scores one is generally considering past performance, what has been accomplished. When analyzing intelligence test scores one is generally looking forward to the future, predicting performance.

It is well to keep in mind the fact that intelligence is inferred from achievement. We have no direct measures of intelligence completely divorced from achievement.

Assumption That Interests and Aptitudes Are Synonymous

Probably few users of standardized tests would acknowledge a belief that interests and aptitudes are the same thing. Webster defines the terms in clearly different domains. Yet how many users of standardized interest inventories can truthfully say that they have never made the jump from a high percentile score in "Persuasive" to the suggestion that Bill probably could succeed in sales activities? To say that Bill seems to be interested in many of the same things that are of interest to people who work in occupations that require influencing other people may be accurate. But to say that Bill will likely succeed in one of these occupations is to make the unwarranted jump from interest to aptitude.

There is evidence that interests and

aptitudes are correlated, but not at a level that allows us to predict one from the other with a high degree of accuracy. This is not to say that interest inventories are useless, but their use might well center on their motivational attributes, on their power to stimulate pupil concern over long range planning.

Misconception of the Meanings of Certain Types of Derived Scores

Students of education have been and are continuing to be taught that an intelligence quotient is obtained by dividing mental age by chronological age, and that mental age is determined by the test performance of students at different age levels. Yet, as a matter of fact, very few of the IQ's to be found in the cumulative folders of elementary and secondary schools today are quotient scores at all. IQ's are standard scores, just as are z scores, T scores, stanines, College Board scores, and others for almost every widely used intelligence test.

It is true that IQ's originally were quotient scores. But, primarily for statistical reasons, the deviation IQ was developed some years ago and has since met with almost universal adoption. Even the Stanford-Binet test switched to a deviation IQ in 1959. The change from a quotient score to a standard score has not necessitated any drastic change in interpretation. Yet it seems to the writer that test users would be well advised to stop paying lip service to a type of score that no longer exists, and to become familiar with standard scores, the type of scores actually being used with our intelligence tests.

The grade placement or grade equivalent score is another type of derived score that is frequently misinterpreted or overused. All too often it is assumed that a grade placement score is an indication of the grade to which a pupil should be assigned. It does not provide that type of information; it simply tells a user whether a pupil is doing high, average, or low quality work. A percentile rank also provides the same assessment of level of work, yet avoids the danger of overinterpretation. If one wishes to compare a pupil's achievement on two different tests in an achievement battery (e.g., reading level versus arithmetic level), a grade equivalent score may lead one to an important misinterpretation. Because of the variability (standard deviation) of grade placement scores from test to test it is possible for a sixth grade pupil to be at the 90th percentile in both reading and arithmetic, yet receive grade placement scores of 8.8 in reading and 8.0 in arithmetic.* If a teacher sees only the grade placement scores of 8.8 and 8.0 he may assume superiority in reading, whereas the two scores represent equivalent performance. For test-by-test comparisons in elementary level achievement batteries percentile scores should be used.

Grading or Promoting Pupils

Standardized achievement tests are designed with certain purposes in mind.

*Iowa Tests of Basic Skills, end-of-year percentile norms for sixth grade pupils.

In general, test authors attempt to identify those skills and understandings that are common to most educational programs. They look for the common denominators; they make no attempt to cover those unique aspects of content that a particular school system may incorporate in its curricular offering. They cannot attempt to reflect a particular teacher's goals for his own pupils. Thus, while achievement test results represent very useful assessment of certain skills and understandings that are common to many classrooms, they should not be used to replace a teacher's own assessment devices.

In many schools standardized achievement tests are given toward the beginning of the school year. They are used to look ahead rather than to look back, to diagnose rather than to evaluate or grade. In those schools that use standardized achievement tests at the end of the year, it may be interesting for a teacher to compare the results with his own judgments. It is not wise for the test results to be used to replace his judgments, in either grading or promotion.

Judging Effectiveness of Instruction

Just as standardized tests are not designed to be used for grading pupils, they are not not designed to be used for grading teachers. Many of the outcomes of classroom instruction cannot be programmed in standardized tests. Those that can be programmed in tests may not be meaningful because of different emphases, different content, and different grade placement in a particular school.

Of special concern is the attitude engendered in teachers in a school attempting to assess instruction through achievement tests. When test results are used to judge teachers, teachers soon learn to teach for the tests.

It is interesting to note that in some instances teachers even feel compelled to "teach for" ability tests. They somehow feel that it isn't respectable to turn in a set of IQ scores for filing in a cumulative record unless all or almost all of them are at least 100. Such a feeling, of course, is based on a misconception of the meaning of intelligence. A teacher may be cutting his own throat with such high scores, for if his pupils all are above average in ability they may be expected to show equally high achievement levels.

Comparing Results from Different Tests

There is a very natural tendency for test users to assume that a language usage score from one test is directly comparable to a language usage score from another test, that an IQ from one test means the same thing as an IQ from another. When making such assumptions one tends to forget two very important characteristics of standardized tests:

1. Test authors do not build their tests on the same specifications, following the same blueprints. Each one develops his own specifications for test construction. There usually is considerable overlap between the

plans for a language usage test developed by one author and the plans developed by another. However, there is never a complete overlap. Scores from two tests measuring the same attribute vary to a certain extent because the test designs vary.

2. The norms for different tests are based upon different groups of pupils. Each test author aims at securing a truly random population of pupils for use in standardizing his test. Each author falls somewhat short of his goal. While it is correct to assume that test norms for two different arithmetic tests are based on groups with considerable overlap in achievement, it is not correct to assume 100-per-cent overlap.

Thus, two IQ's derived from two different intelligence tests are not exactly equivalent. It has been demonstrated that IQ's can vary as much as 5 or 10 points between different tests for no other reason than that they are different tests.

Sometimes one hears this objection: "But how can IQ's be different from different tests? I thought that all good intelligence tests correlated well with each other." It is true that the correlations between different intelligence tests are generally sizable, and many times are almost as high as the reliability of the separate tests. Such correlations do *not* guarantee comparability of norms. Such correlations simply say that pupils taking the different tests will tend to get scores putting them in the same relative rank order but not with the same scores. For example, suppose one were to take a set of IQ's (or any other test score) and add 50 points to each score. The correlation between the original IQ's and the new scores would be perfect, yet the two sets of scores would be 50 points apart.

As was mentioned earlier, we are in the middle of a boom period of testing. If the users of tests do a good job of interpreting the results for the improvement of our understanding of boys and girls and for the improvement of instruction, the boom will level off on a satisfactory plateau of test use. If the users of test results fall into the various misuses, overuses, and misconceptions that are possible, the boom will most certainly be followed by a "bust."

It is the thesis of this article that the consumers of test scores must be thoroughly conversant with proper methods of test use and must studiously avoid misuses, overuses, and misconceptions.

"A little knowledge is a dangerous thing."

6.4 *Improving the Competence of Teachers in Educational Measurement*

ROBERT L. EBEL

Reprinted from the *Clearing House,* 1961, **36,** 7–71, with permission of the author and the *Clearing House.* Robert L. Ebel is a specialist in measurement and educational psychology, who played a leading role in developing tests at the Educational Testing Service and who is now affiliated with the School of Education at Michigan State University.

This paper is included as a correlative to the preceding paper by Womer, who pointed up some of the shortcomings teachers have with respect to test usage. Ebel's article picks up where Womer's leaves off, for he explores the problem further and ends by making some practical suggestions as to how teachers can be helped to improve their skills in using tests.

It is of the utmost importance to educational progress that the competence of teachers to measure educational attainments be improved. Far more harm, perhaps ten times as much harm, is currently being done to student learning as a result of the shortcomings of the classroom tests by which a student's educational efforts are largely stimulated, directed, and evaluated, than is being done by all the faults of external testing programs. Let us concede that some faults do exist in external testing programs—faults of coercion, of duplication, and of misuse of test results. Let us also concede that these faults detract from the great service that external testing programs can render to education. But let us not allow the recognition of this mote to distract our attention from the beam. I am fully persuaded that the current problem in testing which most urgently requires the attention of all professional educators is that of improving the tests made and used by the classroom teacher.

To establish the importance of improving teacher competence in measurement, it is necessary to show not only that measurement of educational attainments is needed and possible but also that teachers are currently deficient in getting this job done. I know of little objective data which would confirm this critical view of teacher competence in measurement. It does, however, seem to be quite generally supported by the opinions of all concerned—measurement specialists, school administrators, students, parents, even by teachers themselves.

What are these deficiencies? What are the most serious errors teachers commonly make in measuring educational attainments? A review of the comments of competent educators who can also claim competence in educational measurement leads to this list of seven prevalent errors.

First, teachers tend to rely too much on their own subjective, but presumably absolute, standards in judging educational attainments. The essential relativity of educational attainments, and the unreliability of subjective judgments, have been demonstrated over and over again. Yet few teachers have been persuaded to use pooled judgments in co-operative test construction and markings or to recognize their inevitable use of relative standards in evaluating student attainments.

Second, teachers tend to put off test preparation to the last minute, and then to do it on a catch-as-catch-can basis. Not only is a last-minute test likely to be a bad test, but also it cannot possibly have the kind of motivating and guiding influence on student study and learning that a test planned, developed, and described to the students early in the course could have.

Third, many teachers use tests which are too poorly planned, too short, or too inefficient in form to sample adequately all the essential knowledge and abilities in the area of educational attainment covered by the tests. Essay tests have many virtues, but neither adequacy of sampling nor reliability of scoring can ordinarily be attributed to them.

Fourth, teachers often place too much emphasis on trivial or ephemeral

details in their tests, to the neglect of basic principles, understandings, and applications. It is, for example, far less important to know the year in which the Articles of Confederation were drawn up than to know what proved to be their basic weaknesses, yet many teachers are inclined to ask about the date rather than about the weaknesses.

Fifth, teachers often write test questions, both essay and objective, whose effectiveness is lowered by ambiguity, or by irrelevant clues to the correct response. Too seldom do they seek even one independent review of their questions by a competent colleague.

Sixth, many teachers overlook, or underestimate, the magnitude of the sampling errors which afflict test scores. If no error has been made in scoring individual responses or in adding these to obtain the test score, they presume that the score is absolutely accurate. Differences as small as one score unit are considered to reflect significant differences in attainment.

Seventh, most teachers fail to test the effectiveness of their tests by even a simple statistical analysis of the results from the test. They do not even calculate a mean, to see if the test was appropriate in difficulty, or a standard deviation, to see if it differentiated well among the students, to say nothing of calculating the reliability of the scores. Yet calculations of these statistics can be quite simple. There is no better way to develop skill in testing than to analyze systematically the results of previous efforts, and to compare the findings of these analyses with ideal standards.

Let us now approach the matter more positively by considering what a teacher needs to know or to be able to do if he is to be competent in measurement. A detailed consideration of all aspects of this kind of competence cannot be presented here, but some of the major elements can be outlined.

A teacher who is competent in educational measurement should:

1. Know the educational uses, as well as the limitations, of educational tests.
2. Know the criteria by which the quality of a test should be judged and how to secure evidence relating to these criteria.
3. Know how to plan a test and write the test questions to be included in it.
4. Know how to select a standardized test that will be effective in a particular situation.
5. Know how to administer a test properly, efficiently, and fairly.
6. Know how to interpret test scores correctly and fully, but with recognition of their limitations.

This constitutes a fairly large order, but not one beyond the capabilities of most teachers. How can we go about developing this kind of competence?

While a great variety of things might be done to foster improvements in teacher competence in measurement, special emphasis may be focused on only three: increased attention to educational measurement in teacher-training programs; provision of special testing services to teachers in school systems; and special organization of in-service training programs in measurement for teachers.

The first suggestion, increased emphasis on educational measurement in programs of teacher preparation, seems so obvious as to call for little elaboration or defense, once the importance of measurement in education and the current deficiencies of teachers are granted. Yet this emphasis will surely be opposed by those outside the profession who seek more stress on content and less on methods, and by those inside the profession who have other axes to grind. If, to make room for a solid course in measurement, it is necessary to eliminate some other professional courses, room can be made. In almost all programs for teacher preparation, some courses are so barren of practical value, so duplicative of the content of other courses, or so concerned with formalizing the obvious, that elimination of them would pain only those professors whose livelihood depends on having the colleges continue to require courses of all students.

Some of my professional colleagues have urged legislative requirement of credit in educational measurement as a condition for certification to teach. I would oppose this suggestion on the ground that such legislative requirement tends to be both unduly restrictive and practically ineffective. But an alternative suggestion which strikes me as highly reasonable is that teacher competence in general, and in special areas such as educational measurement, be judged not only from credit in courses taken but also, in part, from performance on appropriate examinations.

The second suggestion would require a school system to employ a staff member with special competence in testing. The main function of this person would be to generate concern over inadequate educational measurement and to offer assistance in improving them. I know of no other way in which a professionally competent superintendent, serving an educationally alert and sympathetic community, could move more quickly and surely, at less cost, to improve the educational program of his district. Concern with the quality of tests inevitably involves concern with the appropriateness of objectives and curriculums and with the quality of teaching. And concern is an essential prerequisite of change and improvement.

A number of the larger and better school systems have made effective use of such test specialists for years. There appears to be an accelerating trend for other school districts to follow their example. Given reasonable freedom of action and sympathetic administrative support, a good specialist in educational measurement can do many things for the school and the community. He can help them to define *their* educational objectives specifically and meaningfully. He can help staff and students to maintain high motivation. He can help to provide information on the capabilities and achievements of students, information which makes efficient teaching easier for the staff and effective leadership easier for the administration. Above all, he can help teachers improve their competence in the essential task of measuring educational achievement in the classroom.

Schools exist to educate children, but it is the exceptional teacher or

school administrator who can say precisely or meaningfully how much education his pupils are getting. Administrators can discuss per-pupil costs, pupil-teacher ratios, and average daily attendance with specificity and assurance, but not the educational achievement of the pupils. Teachers can speak fluently of objectives and methods, plans and resources, but ordinarily they find it difficult to say definitely how much education has been achieved. If tests can help to supply this information, and indeed they can, what valid excuse is there for our failure to use them more competently?

My final suggestion of a means for improving the competence of teachers in measurement grows out of experience with a variety of in-service programs intended to help teachers solve measurement problems. These programs, variously referred to as conferences, seminars, institutes, or workshops, have ranged from an afternoon lecture to a three-day preschool program with several follow-up meetings later in the year. Some of these programs were sponsored by a single school system and involved the teachers of that school in all subject areas and at all levels. Others reflected the interest of a single professional group, such as engineers or nurses.

Two main weaknesses have been apparent in the in-service training programs of this type with which I have been acquainted. One is their brevity. While an hour or two a year spent in considering measurement problems under the guidance of a specialist is far better than nothing at all, it is unreasonable to suppose that satisfying, enduring progress in solving the manifold problems of educational measurement, or in developing the requisite knowledge, understanding, and skills, can be made in so short a time.

The second weakness of many of these programs is that they involve too much talking and too little doing. For the cultivation of a practical art like educational measurement, sound pedagogy requires a mingling of theory and practice. A means by which theory and practice can be combined conveniently and effectively is what I would propose.

During the past four years I have watched hundreds of skilled teachers at work on committees charged with the development of new tests for the College Entrance Examination Board and for the Educational Testing Service. (There have been thousands of others whose work I was not able to observe.) Over and over, from the teachers themselves and from our own staff members, have come testimonies as to the educative values of the experience. I am strongly persuaded that the best way to learn how to make good educational tests and how to use them skillfully is to cooperate with other teachers, under expert supervision, in the construction and use of some important tests.

Something along these lines is what I would propose as the ideal program of in-service training for improving the competence of teachers in measurement. Suppose that a school administrator and his staff have decided to focus attention for a year on the improvement of classroom testing. Suppose they engage a specialist in educational testing to meet with them five times during the year, at intervals of

six weeks or so, for a day or two. Participation in the initial program might well be limited to five, six, or seven groups of four to six teachers each.

The goal of each group would be to make, to use, and to analyze a quality test in a subject which all members of the particular group were teaching. Examples of the subject areas in which these tests might be developed are: fourth-grade mathematics; sixth-grade geography; eighth-grade English; or high-school history, chemistry, or economics.

The first meeting of each participating group would be devoted to a description of the entire project, with special consideration of the first step —the preparation of specifications for the test to be developed. Sample specifications would be presented for study and analysis. Between the first and second sessions each teacher group would work out the specifications for its test. These could be reviewed at the second meeting, and work on item writing would be launched. Third meeting could be devoted to item review and test assembly, the fourth to test administration and analysis, and the fifth to a review of the test developed and of the entire project as a learning experience.

While I have never participated in a project exactly like this, the most successful ones I know about have much in common with it. I have no doubt that it would produce not only a handful of excellent tests but also a sizable group of teachers whose competence in measurement was vastly improved and, by current standards, highly respectable.

Don Carl Steffen from Rapho Guillumette

PART 7

Individual Differences: Intelligence and Creativity

7.1 Heredity and Environment: A Controversy Over IQ and Scholastic Achievement

ARTHUR R. JENSEN

Reprinted from *Berkeley Centennial Fund, University of California,* 1969, **2**(4), 4–6, with permission of the author and the University of California at Berkeley. Arthur R. Jensen is a member of the educational psychology faculty and is also on the staff of the Institute of Human Learning of the University of California at Berkeley.

In this article, Arthur R. Jensen summarizes the main points from his longer paper, "Intelligence, scholastic performance, and heredity," that appeared in the *Harvard Educational Review,* Winter, 1969. Psychologists have long maintained that although individual differences in IQ are to some degree genetically determined, IQ differences between racial, social-class, and cultural groups are the product of different environments. Jensen's statement that differences between racial groups are genetically, rather than environmentally, determined, has hence been widely criticized by behavioral scientists and laymen alike. The present article recapitulates some of his arguments, and the article by Anastasiow that follows offers a rebuttal.

Compensatory Education

In my article, I first reviewed the conclusion of a nationwide survey and evaluation of the large, Federally funded compensatory education programs done by the U.S. Commission on Civil Rights, which concluded that these special programs had produced no significant improvement in the measured intelligence or scholastic performance of the disadvantaged children whose educational achievements they were specifically intended to raise. The evidence presented by the Civil Rights Commission suggests to me that merely applying more of the same approach to compensatory education on a larger scale is not likely to lead to the desired results, namely increasing the benefits of public education to the disadvantaged. The well-documented fruitlessness of these well-intentioned compensatory programs indicates the importance of now questioning the assumptions, theories, and practices on which they were based. I point out, also, that some small-scale experimental intervention programs have shown more promise of beneficial results.

I do *not* advocate abandoning efforts to improve the education of the disadvantaged. I urge increased emphasis on these efforts, in the spirit of experimentation, expanding the diversity of approaches and improving the rigor of evaluation in order to boost our chances of discovering the methods that will work best.

The Nature of Intelligence

In my article, I pointed out that IQ tests evolved to predict scholastic performance in largely European and North American middle-class populations around the turn of the century. They evolved to measure those abilities most relevant to the curriculum and type of instruction, which in turn were shaped by the pattern of abilities of the children the schools were then intended to serve.

IQ or abstract reasoning ability is thus a selection of just one portion of the total spectrum of human mental abilities. This aspect of mental abilities measured by IQ tests is important to our society, but is obviously not the only set of educationally or occupationally relevant abilities. Other mental abilities have not yet been adequately measured; their distributions in various segments of the population have not been adequately determined; and their educational relevance has not been fully explored.

I believe a much broader assessment of the spectrum of abilities and potentials, and the investigation of their utilization for educational achievement, will be an essential aspect of improving the education of children regarded as disadvantaged.

Inheritance of Intelligence

Much of my paper was a review of the methods and evidence that lead me to the conclusion that individual differences in intelligence, that is, IQ, are predominantly attributable to genetic differences, with environmental factors contributing a minor portion of the variance among individuals. The heritability of the IQ—that is, the percentage of individual differences variance at-

tributable to genetic factors—comes out to about 80 per cent, the average value obtained from all relevant studies now reported.

These estimates of heritability are based on tests administered to European and North American populations and cannot properly be generalized to other populations. I believe we need similar heritability studies in minority populations if we are to increase our understanding of what our tests measure in these populations and how these abilities can be most effectively used in the educational process.

Social Class Differences

Although the full range of IQ and other abilities is found among children in every socioeconomic stratum in our population, it is well established that IQ differs on the average among children from different social class backgrounds. The evidence, some of which I referred to in my article, indicates to me that some of this IQ difference is attributable to environmental differences and some of it is attributable to genetic differences between social classes—largely as a result of differential selection of the parent generations for different patterns of ability.

I have not yet met or read a modern geneticist who disputes this interpretation of the evidence. In the view of geneticist C. O. Carter: "Sociologists who doubt this show more ingenuity than judgment." At least three prominent sociologists who are students of this problem—Sorokin, Bruce Eckland, and Otis Dudley Duncan—all agree that selective factors in social mobility and assortative mating have resulted in a genetic component in social class intelligence differences. As Eckland points out, this conclusion holds *within* socially defined racial groups but cannot properly be generalized *between* racial groups, since barriers to upward mobility have undoubtedly been quite different for various racial groups.

Race Differences

I have always advocated dealing with persons as individuals, each in terms of his own merits and characteristics and am opposed to according treatment to persons solely on the basis of their race, color, national origin, or social class background. But I am also opposed to ignoring the causes of the well-established differences among racial groups in the distribution of educationally relevant traits, particularly IQ.

I believe that the causes of observed differences in IQ and scholastic performance among different ethnic groups is, scientifically, still an open question, an important question, and a researchable one. I believe that official statements, such as "It is a demonstrable fact that the talent pool in any one ethnic group is substantially the same as in any other ethnic group" (U.S. Office of Education, 1966), and "Intelligence potential is distributed among Negro infants in the same proportion and pattern as among Icelanders or Chinese, or any other group" (U.S. Dept. of Labor, 1965), are without scientific merit. They lack any factual basis and must be regarded only as hypotheses.

It would require more space than I

am allotted to describe the personal and professional consequences of challenging this prevailing hypothesis of genetic equality by suggesting alternative hypotheses that invoke genetic as well as environmental factors as being among the causes of the observed differences in patterns of mental ability among racial groups.

The fact that different racial groups in this country have widely separated geographic origins and have had quite different histories which have subjected them to different selective social and economic pressures make it highly likely that their gene pools differ for some genetically conditioned behavioral characteristics, including intelligence, or abstract reasoning ability. Nearly every anatomical, physiological and biochemical system investigated shows racial differences. Why should the brain be any exception? The reasonableness of the hypothesis that there are racial differences in genetically conditioned behavioral characteristics, including mental abilities, is not confined to the poorly informed, but has been expressed in writings and public statements by such eminent geneticists as K. Mather, C. D. Darlington, R. A. Fisher, and Francis Crick, to name a few.

In my article, I indicated several lines of evidence which support my assertion that a genetic hypothesis is not unwarranted. The fact that we still have only inconclusive conclusions with respect to this hypothesis does not mean that the opposite of the hypothesis is true. Yet some social scientists speak as if this were the case and have even publicly censured me for suggesting an alternative to purely environmental hypotheses of intelligence differences. Scientific investigation proceeds most effectively by means of what Platt has called "strong inference," pitting alternative hypotheses that lead to different predictions against one another and then putting the predictions to an empirical test.

Dysgenic Trends

More important than the issue of racial differences *per se* is the probability, explicated in my article, of dysgenic trends in our urban slums, as suggested by census data showing markedly higher birth rates among the poorest segments of the Negro population than among successful, middle-class Negroes. This social class differential in birthrate appears to be much greater in the Negro than in the white population. That is, the educationally and occupationally least able among Negroes have a higher reproductive rate than their white counterparts, and the most able segment of the Negro population has a lower reproductive rate than its white counterpart.

If social class intelligence differences within the Negro population have a genetic component, as in the white population, the condition I have described could create and widen the genetic intelligence differences between Negroes and whites. The social and educational implications of this trend, if it exists and persists, are enormous. The problem obviously deserves thorough investigation by social scientists and geneticists and should not be ignored or superficially dismissed as a result of

well-meaning wishful thinking. The possible consequences of our failure seriously to study these questions may well be viewed by future generations as our society's greatest injustice to Negro Americans.

Learning Ability and IQ

The article also dealt with my theory of two broad categories of mental abilities, which I call intelligence (or abstract reasoning ability) and associative learning ability. These types of ability appear to be distributed differently in various social classes and racial groups. While large racial and social class differences are found for intelligence, there are practically negligible differences among these groups in associative learning abilities, such as memory span and serial and paired-associate rote learning.

Research should be directed at delineating still other types of abilities and at discovering how the particular strengths in each individual's *pattern* of abilities can be most effectively brought to bear on school learning and on the attainment of occupational skills. By pursuing this path, I believe we can discover the means by which the reality of individual differences need not mean educational rewards for some children and utter frustration and defeat for others.

Conclusion

Without a doubt, my article has provoked serious thought and discussion among leaders in genetics, psychology, sociology, and education concerned with these important fundamental issues and their implications for public education. I expect that my work will stimulate further relevant research as well as efforts to apply the knowledge gained thereby to educationally and socially beneficial purposes.

In my view, society will benefit most if scientists and educators treat these problems in the spirit of scientific inquiry rather than as a battlefield upon which one or another preordained ideology may seemingly triumph.

7.2 Educational Relevance and Jensen's Conclusions

NICHOLAS ANASTASIOW

Reprinted from the *Phi Delta Kappan,* 1969, **51**, 32–35, with permission of the author and Phi Delta Kappa, Inc. Nicholas Anastasiow is director of the Institute for Child Study, Indiana University, and was formerly program director of the Education Improvement Project of Durham, North Carolina, working with poverty children between the ages of two and nine.

In the preceding article, Jensen set forth his position that intelligence differences between racial groups are genetically determined. In the present article, Anastasiow cites the critics whose statements were published in the same issue of the *Harvard Educational Review* that carried Jensen's original article, and adds further data from his own work with poverty children to refute Jensen's thesis.

Probably nothing on today's educational scene has created more attention and controversy than Arthur R. Jensen's *Harvard Educational Review* article of last winter and the discussions by Jerome Kagan, J. McV. Hunt, J. F. Crow, Carl Bereiter, David Elkind, Lee Cronbach, and W. F. Brazziel published in the spring issue. National news media gave the original article wide coverage, with focus on specific sections, stimulating a wide variety of continuing reaction. Diverse though the responses may be, remarkably similar issues keep cropping up. I shall review these recurrent themes.

It is generally agreed that Jensen's article exhibits breadth of scholarship, that it makes a contribution to scholarly dialogue, and that there is a genetic factor in intelligence. There has been general disagreement about two of Jensen's conclusions, those two which have been most widely quoted:

1. "That genetic factors are strongly implicated in the average Negro-white intelligence difference. The preponderance of evidence is, in my [Jensen's] opinion, less consistent with a strictly environmental hypothesis than with a genetic hypothesis, which, of course, does not exclude the influence of environment or its interrelation with genetic factors." [1]

2. [That] compensatory education has been tried and it apparently has failed." [2]

Several of the *Review* discussants have also raised questions as to conclusions Jensen draws from the data of his research and that of other research he reports, and there has been criticism of his failure to include consideration of the findings of other pertinent studies, particularly in the area of socialization and early child growth and development.

In the main, Jensen discusses intelligence, how it is measured, its correlates, and its hereditability. He reviews a wide range of research studies in genetics, intelligence testing, and psychology, and draws the conclusion that attempts to raise intelligence are fruitless. He concludes not only that there are genetic differences between individuals but that there are genetic differences in intelligence between whites and Negroes. Further, he says that specific aptitudes should be identified and school curriculum and techniques should be designed to teach to these aptitudes.

Jensen argues that the "environmentalists" have misled us in regard to how much environment can modify intelligence. He bases his argument on a series of earlier research studies in which he and his students worked extensively with white, Negro, and Mexican-American students. These studies appear to be well executed and controlled. They suggest that, while there are social class differences in children's performances of complex tasks, there do not seem to be the same kinds of differences among social classes on associative learning tasks.

In general, the discussants in the spring issue of the *Review* agree with the need to recognize special abilities and aptitudes, and consider Jensen's scholarship of high caliber. However, they do not all agree when Jensen attributes the difference in intellectual functions of the lower-class child (par-

ticularly the Negro) not to the psycho-socio-cultural deprivation hypothesis so ably prepared by Hunt[3] but to basic genetic factors. Jensen suggests that about 80 percent of intelligence can be accounted for by genetics alone. Crow appears to agree. Cronbach[4] raises some question about this figure. And there are newer conceptions of genetics not discussed by Jensen which would set this limit nearer 50–60 percent.

What is troubling about Jensen's hypothesis of racial differences in I.Q. is that he appears to seek truth and make known his findings, yet seems to close off alternative hypotheses to explain the data. At the same time, he asks us to consider the tenability of an older hypothesis of racial differences in I.Q. which his data, to this reviewer, fail to support. Equally ironic is the fact that some of his severest critics in the news media appear to be closed to the hypothesis he suggests. Rather than analyze data and offer alternative interpretations, they condemn Jensen. Neither approach will serve science and education.

To support his hypothesis of racial differences in I.Q., Jensen draws upon data from several areas, one of which consists of studies of genetic influence on height and weight. To Jensen, these suggest racial and genetic limitations. As Cronbach indicates in his discussion, Jensen is on weak ground here. Height and weight can be modified and are not under purely genetic control, as an earlier study by Greulich[5] of Japanese-Americans and their Japanese cousins strongly documents.

Greulich's studies controlled for genetic factors and found marked differences in height, weight, and chest expansion in favor of the California subjects who came from the same gene pool as the comparison group, but whose dietetic, climatic, and other environmental conditions were more conducive to optimum development. Thus supposed racial traits proved modifiable by environmental conditions.

Jensen dismisses the environmental-influence-on-intelligence point of view rather lightly, although he does mention some studies. He questions it by stating that "disadvantaged children are not reared in anything like the degree of sensory and motor deprivation that characterizes, say, the children of the Skeels study."[6] *However,* the breadth of studies by Spitz, Bowlby, Anna Freud, Skodak and Skeels, and Skeels and Heinicke[7] on the effects of deprivation, separation, and adoption have well demonstrated the effects of physical, emotional, perceptual, and environmental conditions on intellectual functioning and adult academic attainments.

Perhaps, in working with 5-year-olds in California, Jensen has not seen the effects of lack of stimulation on the 2- to 5-month-old infant. There is, however, ample evidence reported by reputable research psychologists to support the fact that there are marked conditions of economic and social and psychological lacks among the poor of this nation.

The issue of environment versus heredity was ably handled by Anastasi and Foley[8] years ago as an interaction phenomenon. As Piaget and Inhelder[9] suggest, mental development is influenced by maturation or organic growth,

the role of exercise and acquired experience, social interaction and transmission, and an internal mechanism of equilibration. Knowledge, to Piaget and Inhelder, is derived from action; thus the culture, the socialization processes and techniques used to train the child, greatly influence what we measure on an I.Q. test. If Jensen's otherwise excellent article has a major weakness, it is in its naive conception of growth and development and the critical area of the socialization process on the intellectual functioning of the child. For example, both physical and language development appear globally undifferentiated in the young child and are progressively differentiated through learning and experience. As Elkind[10] states in his review, the Piagetian view is that intelligence develops through experience.

The paradox of Jensen's article is that it contains one of the clearest descriptions of what is required for school success as schools now exist, then states a faulty assumption with respect to the way children acquire skills that lead to school success. Let's look at Jensen's excellent paragraph:

Our thinking almost always takes as granted such features as beginning formal instruction at the same age for all children (universally between ages five and six), instruction of children in groups, keeping the same groups together in lock-step fashion through the first several years of schooling, and an active-passive, showing-seeing, telling-listening relationship between teacher and pupils. Satisfactory learning occurs under these conditions only when children come to school with certain prerequisite abilities and skills: an attention span long enough to encompass the teacher's utterances and demonstrations, the ability to comprehend verbal utterances and to grasp relationships between things and their symbolic representations, the ability to inhibit large-muscle activity and engage in covert "mental" activity, to repeat instruction to oneself, to persist in a task until a self-determined standard is attained —in short, the ability to engage in what might be called self-instructional activities, without which group instruction alone remains ineffectual.[11]

Children of middle-class homes have been prepared to meet these conditions and are carefully trained to be able to function within the school situation, as shown by Sears, Maccoby and Levin, Loevinger, and Schaeffer.[12] In the majority of cases, children of lower-class homes have not received such training (see Hess and Shipman[13]). When the lower-class child begins school, he is usually required to act as if he has already mastered the necessary prerequisites for verbally oriented group instruction, though in fact he has not because the socialization process—particularly of the Negro—has not so prepared him. Hess and Shipman, Anastasi, Hertzig, Birch, Thomas and Mendez,[14] and others have shown in comparisons of lower-class mothers with middle-class mothers marked differences in what is taught and how it is taught. What is taught by the lower-class mother appears to be nonverbal cognitive systems which may reflect intellectual functionings that our typical intelligence tests do not measure.

However, as McNeill[15] states, in our schools verbal facility is usually measured by the most peripheral aspects of language; phonology and morphology are taken as signs of intelligence. Kagan[16] also suggests that the deprived or poverty child has not had the necessary early stimulation which with the middle-class child begins as early as 3–6 months.

The issue of Negro-white intellectual differences is often argued without taking into account those very early child trainings which Piaget[17] and Hunt[18] regard as crucial. Isolated from the main stream of America, many poverty mothers have not been provided with the necessary techniques to maximize their children's intellectual potential.

The Negro is frequently compared disparagingly, as Jensen infers, with European immigrants who have been assimilated into the culture and are functioning successfully within it, acquiring power and prestige positions. The analogy is a gross fallacy, because it fails to take account of two relevant facts: 1) European immigrants were usually motivated toward the same kind of success as white Americans and brought with them a predominantly verbally oriented culture. 2) The immigrant's child, once he mastered the outward manifestations of middle-class acceptability in speech and dress, could not be stigmatized as to ethnic origin and had access to the middle- and upper-class prestige positions. No matter how accurate a Negro's speech, how skillfully he masters the formalities of middle-class attitudinal and value systems, he is still unable to melt into the culture because of his color. Unfortunately, teaching to individual differences will not assure adult success until we cast aside the Neanderthal notion that skin color is related to intelligence. Guskin's[19] recent work is a case in point. Teachers listening to tape recordings of speech by Negro children purportedly reading their own compositions rated them lower than white children who read the same passages.

Jensen does, however, present some very reasonable suggestions for beginning instruction with poverty children. He reviews his own research and that of Lesser, Fifer, and Clark[20] to demonstrate different patterns of ability among ethnic groups which may not be related to social class. If the deprived child uses different modes of thought, we should teach to those modes, Jensen suggests. There is a difference, however, between Jensen's position that there is a limited number of skills that can be taught to poverty children and conclusions that can be drawn from research with lower-class children. Jensen feels that most school skills can be taught and acquired by associative learning, which he calls Level I learning. He appears to imply that racial lower-class groups are so different from others as to preclude their developing abstract reasoning and conceptual thinking (Level II learning). Apparently, he believes that the lower-class minority has failed with these tasks and cannot be expected to master them. Other learning theorists postulate that styles or modes of thought do not preclude problem solving. Rather,

they believe that how a child is taught abstract reasoning and problem-solving skills should vary, depending upon the child's strengths or skill profiles. Jensen seems to want to close the door that it has taken a decade of special educators and modern curriculum personnel in the Dewey,[21] Taba[22] tradition to open. That is, the major question is not why a child can't function now but how we can provide him with an educational program so that he can. Therefore, although Jensen agrees that early education programs have not focused on the relevant tasks, his conclusion that they have failed and will fail does not logically follow from his discussion. Both Hunt[23] and Kagan[24] point out that early education programs have not been developed adequately for assessment to be valid at this time.

Many compensatory education programs have been based upon what has been successful with the "task-oriented" middle-class child. We are only beginning to find out what are appropriate interventions for the deprived child. In addition, the compensatory programs have been built too often upon the weakest of the middle-class preschool models, which deal largely with drill and practice techniques and total group activities rather than individual child opportunities to explore, manipulate, and deal with the environment. Maturation of functions demands different modes of transaction with the child, and learning requires reward, reinforcement, and tasks near the child's current level of functioning.[25] Our teaching procedures must come to match the competencies and the level of development the child brings with the requirements of the task to be learned. To do this takes time and careful analysis of the learning task and the stage of the child's development. As Tyler[26] long ago pointed out, it often takes teachers three years to be able to perform the teaching procedures that an innovative program intends to implement.

If there are weakness in these programs (and there undoubtedly are), they may well reside in our techniques of training teachers to meet marked individual and subcultural differences and our lack of stronger in-service training programs. Until we are more successful in preparing teachers, we will not identify those components of instruction that enable the child to develop the multiplicity of intellectual competencies available at birth regardless of what the ultimate capacity may be.

Jensen could have assisted us greatly by establishing what the real issue for educators is. It is *not* one of heredity versus environment; it is concerned with discovering what kind of environmental stimulations are necessary to reach the potential of what is inherited. We must adopt a more dynamic approach to intelligence and mental development than that presented by Jensen.

Piaget and Inhelder[27] state it beautifully: "It may even seem that effective dynamic factors provide the key to all mental development and that in the last analysis it is the need to grow, to assert oneself, to look, to be ad-

mired that constitutes the motive force of intelligence, as well as behavior in its totality and its increasing complexity."

References

1. Arthur R. Jensen, "How Much Can We Boost I.Q. and Scholastic Achievement?," *Harvard Educational Review.* Winter, 1969, p. 82.
2. *Ibid.,* p. 2.
3. J. McVicker Hunt, *Intelligence and Experience.* New York: Ronald Pres, 1961.
4. Lee Cronbach, "Heredity, Environment, and Educational Policy," *Harvard Educational Review,* Spring, 1969, pp. 338–47.
5. William W. Greulich, "A Comparison of the Physical Growth and Development of American-born and Native Japanese Children," *American Journal of Physical Anthropology,* December, 1957, pp. 489–516.
6. Jensen, *op. cit.,* p. 61.
7. [a] Rene Spitz, "Hospitalism: An Inquiry into the Genesis of Psychiatric Conditions in Early Childhood," *Psychoanalytic Studies of the Child.* New York: International Universities Press, 1945, p. 53–74.
   [b] John Bowlby, Mary Arnsworth, Mary Boston, and Dina Rosenbluth, "The Effects of Mother-Child Separation," *British Journal of Medical Psychology,* Part 4, 1956, pp. 211–47.
   [c] Anna Freud, *Infants Without Families.* New York: International University Press, 1944.
   [d] Marie Skodak and Harold M. Skeels. "A Final Follow-up Study of One Hundred Adopted Children," *Journal of Genetic Psychology,* September, 1949, pp. 85–125.
   [e] Harold M. Skeels, *Adult Status of Children with Contrasting Early Life Experiences: A Follow-up Study.* Monograph of the Society for Research in Child Development, Serial No. 105, 1966.
   [f] Christoph M. Heinicke, "Some Effects of Separating Two-Year-Old Children from Their Parents," *Human Relations,* May, 1956, pp. 106–76.
8. Anne Anastasi, and John P. Foley, "Proposed Reorientation in the Heredity-Environment Controversy," *Psychological Review,* May, 1948, pp. 23–949.
9. Jean Piaget and Barbel Inhelder, *The Psychology of the Child.* New York: Basic Books, 1969.
10. David Elkind, "Piagetian and Psychometric Conceptions of Intelligence," *Harvard Educational Review,* Spring, 1969, pp. 319–37.
11. Jensen, *op. cit., p.* 10.
12. [a] Robert R. Sears, Eleanor E. Maccoby, and Harry Levin, *Patterns of Child Rearing.* Evanston, Ill.: Row, Peterson, 1957.
    [b] Jane Loevinger, "On the Proportional Contributions of Differences in Nature and Nurture to Differences in Intelligence," *Psychological Bulletin,* December, 1943, pp. 725–56.
    [c] Earl S. Schaeffer, "A Circumplex Model for Maternal Behavior," *Journal of Abnormal Social Psychology,* Volume 59, 1959, pp. 226–36.
13. Robert D. Hess and Virginia C. Shipman, "Early Experiences and the Socialization of Cognitive Modes in Children," *Child Development,* December, 1965, pp. 869–86.
14. [a] *Ibid*
    [b] Anne Anastasi, *Differential Psychology,* 3rd ed. New York: Macmillan, 1958.
    [c] Margaret E. Hertzig, Herbert Birch, Alexander Thomas, and Olga Aran Mendez. *Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands.* Monograph of the Society for Research in Child Development, Serial No. 117, 1968.
15. David McNeill, "The Development of Language," ERIC, Document No. ED021218, 1968.

16. Jerome Kagan, "Inadequate Evidence and Illogical Conclusions," *Harvard Educational Review.* Spring, 1969, pp. 224–77.
17. Jean Piaget, *The Origins of Intelligence in Children* (trans. by Margaret Cook). New York: International University Press, 1952.
18. Hunt, *op. cit.*
19. Judith Guskin, "Current Approaches to the Study of Language Behavior and Some Recent Research on Mental Retardation," mimeographed, 1968.
20. Gerald Lesser, Gordon Fifer, and Donald Clark. *Mental Abilities of Children from Different Social-class and Cultural Groups.* Monograph of the Society for Research in Child Development, Serial No. 102, 1965.
21. John Dewey. *How We Think,* rev. ed. Boston: Heath, 1933.
22. Hilda Taba. *Teaching Strategies and Cognitive Functioning in Elementary School Children,* Cooperative Research Project No. 2404. San Francisco: San Francisco State College, 1966.
23. J. McVicker Hunt. "Has Compensatory Education Failed? Has It Been Attempted?," *Harvard Educational Review,* Spring, 1969. pp. 278–300.
24. Kagan, *op. cit.*
25. Robert M. Gagné, "The Acquisition of Knowledge," *Psychological Review,* July 1962. pp. 355–65.
26. Ralph Tyler. *Appraising and Recording Students' Progress.* New York: Harper & Brothers, 1942.
27. Piaget and Inhelder, *op. cit.*

7.3 *Diversity*

JOHN R. PLATT

Reprinted in abridged form from *Science,* 1966, **154,** 1132–1139, with permission of the author and *Science.* Copyright © 1966 by the American Association for the Advancement of Science. John R. Platt is a biophysicist and associate director of the Mental Health Research Institute at the University of Michigan.

A teacher cannot simultaneously respond in an individually different manner to all students in his class. Much of the time he has to deal with them as a group, treating them as if they were all alike. The fact that the students in a given class are similar in certain ways gives such treatment a degree of validity, although "treating students all alike" is often more a matter of convenience to the teacher than anything else. The great danger, of course, is that teachers who habitually treat all students alike never get around to dealing with many of them on individual terms. As a result, far too much of the teaching-learning process becomes standardized and routinized.

In this article, Platt argues for more freedom and greater variety in the way educational goals are approached. Although the target of much of his criticism is the university, what he says is equally relevant for secondary and even elementary education. Diversity in educational experiences and growth in creativity are closely related. If creativity is a matter of concern to teachers, there must be a great deal more individuality in educational strategies and programs.

I celebrate diversity. Our research, our lives, our goals, our pursuit of excellence are all too homogeneous. La Rochefoucauld writes: "God has put as differing talents in man as trees in Nature: and each talent, like each tree, has its own special character and aspect. . . . The finest pear tree in the world cannot produce the most ordinary apple, and the most splendid talent cannot duplicate the effect of the homeliest skill."

I think he means that other men are not like him in being able to produce maxims of this kind. But what he says is true. How many of us have gotten D's and F's in apple-tree courses simply because the teacher was too narrow to see that we had to be nurtured as pear trees? Progress would be faster and life would be more interesting if we pursued more diverse goals, goals of excellence to be sure, but goals of our own, different from what everybody else is pursuing—and if we tolerated and encouraged the same sort of individuality in others. I want life to be various. I want to see around me not only apple trees but pear trees, not only fruit trees but slow-growing oaks and evergreen pines and rosebushes and bitter but salubrious herbs and casual dandelions and good old spread-out grass. Let us be different, and enjoy the differences!

Diversity in Education

The area where there is perhaps the greatest need of all for more diversity today is the area of education. Students nowadays can hardly realize how much the alternatives available to them have been closed up by the zealous professionalism of the professors in the last 30 years. In the 1930's the colleges knew they had been liberalized by John Dewey and they offered what is now sneered at as a "cafeteria system" of education. Yet what an enriched program it permitted us! When I was an undergraduate physics major at Northwestern University I not only took physics and math courses but I had time for electives that included 2 years of French and 3 years of German (Goethe and Schiller), plus astronomy, economics, philosophy, public speaking, music, and a seminar on the origins of war.

Our present survey courses are more thorough and systematic but not so well tailored to each individual's curiosity and enthusiasm. Many colleges have pushed electives almost out of the curriculum, in favor of so-called "honors programs." All too often these should be called "narrows programs," for what they make is one-dimensional men.

It particularly worries me that physics and chemistry majors and other science majors have now lost most of their free electives. Scientists are now rising to executive positions in business and industry and are becoming advisors on major international and military matters. About one-third of all physicists eventually become administrators. I do not want—and I do not think any sane person wants—a world in which the major decisions on technological and military and international affairs are made by one-dimensional men, men who have never had time to explore art or music or history or phil-

osophy or literature or the nontechnical achievements of mankind!

The only thing that saves us is the fact that the good students learn many things outside the curriculum. I think that in many cases the reputation of the hard-driving schools, both the high schools and the colleges, is not due to the courses or the staff at all, but is due to the quality of the students they are able to get. If you have hot-shots, it makes little difference what you teach them—or whether you teach them at all; they will find out from each other (as the whole human race did!) how to be great contributors to society. The importance of this initial student selection factor has never been sorted out in assessing our schools. Many a school has good graduates not because its education is good but because its students were good when they came in and have not been much damaged.

Even so, the hot-shot dimension is not the only one to be emphasized. Why should we assume or insist that our students have only one important coordinate of variation? This is the fallacy of exams and I.Q. tests.

It is good that Jacob Getzels and Philip Jackson and others have emphasized recently that there is a dimension of "creativity" in students that has little relation to I.Q. How many other such dimensions of achievement are still to be explored?

We do not even allow for the physiological variations in students. Students, like professors, are not all wakeful or sleepy at the same time. We often start by trying to teach them things when they—and we—are half-asleep; and then we try to get them to go to sleep when they are wide awake. Would it be impossible to have classes at one time of day for the skylarks and at another time for nightingales? Even professors might like it. Some of the world's greatest leaders napped in the daytime and worked around the clock. Classes in the evening might lead to the best discussions of all if you could sleep in the morning. I have never understood why these possibilities are not seriously examined by educators, who are supposed to know something about the psychology and physiology of learning.

While we are speaking of the right to physiological diversity, let us not forget the right of some of the students to be women. It is easy to show that prejudices and handicaps to women's education still abound. Fathers send sons to college rather than daughters; and not only fathers but deans will cut off a college girl's financial support if she gets married, where they would not cut off a boy's. I have known professors in several departments who refused to take girls as graduate students on the grounds that they would probably get married and not use the education. The "nepotism rules" of many schools result in failure to hire good women teachers if they have the misfortune to be married to good men teachers, so the image of the woman intellectual that the student sees is almost always that of a woman who has renounced marriage. One great university lost a great woman scientist in this way, through refusing to pay her a salary separate from her husband's—until she became famous.

What is worse, however, is the fact that the colleges and counselors do nothing to combat the double standard of the college men, who may learn far-out things in biology or anthropology but are never shaken out of their conventional station-wagon images of what marriage should be. They go on assuming that the college wife, or the graduate-student wife, is the one who shops and cooks and cleans, even if she is carrying courses and trying to do equal work. The result of this conventional image—which the girls have often picked up as well as the men—is that American women are concentrating on conventional and subordinate jobs and that, compared to women of other countries, they are making fewer and fewer contributions to our national life, either as educators or editors or scientists or doctors or lawyers or judges or legislators or political leaders. We are only getting half-power out of our educated and intellectual women, and it impoverishes us all.

Poverty, Austerity, and Overwork

To come back to the narrowing pressures on student life in general, I think it is not at all clear that the intellectual and the economic pressures on students today are either good education or good economics. Students are probably the most overworked and underpaid class in our society. Their training has now been shown by many studies to be the most important element in the economic development and prosperity of a country, and yet they are not paid as well as their brothers who became plumbers' apprentices. The 18-year-old brother or sister who works in a factory or a store gets off at 5 o'clock and has enough income to have an apartment and a car and books and records and recreation and a paid vacation. He can have guests in and can come in or go out at any hour. But the student is treated, not like his brothers or parents or teachers, but like a monk with a vow of poverty, austerity, and overwork—a vow which is not even his own vow but has been taken for him. He often works until midnight or later at subjects his brothers might never master, and he is supposed to get money from his family, or borrow it, or be grateful for a fellowship that still leaves him below the poverty level. He is frequently locked in at night and forbidden to have a car or an apartment, and has little money for his own books or for good meals or concerts. He is given cafeteria fare in cinder-block buildings and never learns to live like a human being. It is an affluent-society parody of medieval monasticism, with the universities—the primary sources of new economic development today—treated as priestly beggars, and with the professors themselves, who have grown up in the system, approving this treatment of the students and feeling, always, that they have too much money and do not work hard enough.

It is an odd 4-year gap in our economic scheme. Students are overworked and underpaid undoubtedly because they are the only group in our society who are too old for child labor laws to protect them and too young to have the support of a union or of professional-market competition

—as their parents and their professors have—to help them get more civilized hours and treatment.

And, oh, how long are those hours that we are forcing on ambitious students in good high schools and colleges today! You professors who have measured the rates of learning, have you measured the optimum number of hours for intellectual work? Do they agree with the standard homework assignment? It is estimated that a medical student is expected to learn 30,000 bits of information in his first year, or 100 bits per day, if he obeys every demand of the instructors. Is it actually possible to learn at this rate, or does this not simply overload the brain and block any real organization of the material? No wonder the dropout and failure rates are high. No wonder the suicide rate is high.

Men do not become wise and full by studying 14 hours a day, or 10 hours a day, or possibly even 8 hours a day. This is not education for the good life or the good society. There is a limit to human capacity to pack in new knowledge just as there is a limit to the capacity of a stuffed goose. The limit may be no more than a few hours before we need a change of pace for the rest of the day—a period of exercise or recreation or idleness, eating and chatting—if we are really going to assimilate new information and fit it together.

The Narrow Faculties

The trouble is that the faculty itself still thinks this is the only way of education. The student is not taught how to be broad and human because the faculty frequently does not know how to be broad and human. *Nemo dat quod non habet.* No one can give what he does not have. The student is overloaded with information because the professor is overloaded with information, with a piled-up desk and a bulging briefcase. He does not know how to handle it himself, so he passes it on. And many a professor equates education with judgments and grades. I have heard of one man, a kind man in his personal life, who gave out seven F's in a class of 25 undergraduate majors because some students either were not prepared for his 3-hour course or were unwilling to spend 20 hours a week on it, and because he had not the perception or the humanity to tell them earlier that they should not be in the course. This little piece of righteousness will cost these unfortunates hundreds of thousands of dollars in lost fellowships and graduate education and potential job opportunities over their lifetimes. In any other line of work, a man who did such a thing could be sued. In a university, he tells his colleagues it shows how poor the students are today, and they cluck sympathetically. Sometimes such men mellow as they mature, but all too often these black-and-white academics only get more and more self-righteous all the way to retirement.

The student comes for teaching and what he gets is grades. We are hypnotized by grades. They seem so exact and discussable. I have seen departments where one-quarter of the teachers' time and energy was spent in making up exams and grading them. If

any administration doubts this, let it measure the ratio. This amount of time spent with individual students could have pulled many of them over the borderline; but we prefer to retreat to written questions. It gives us renewed proof that our students are one-dimensional. What Montessori said should be written on every bluebook in letters of fire: "The business of a teacher is to teach, not to judge." The business of a professor is to give, not grades, but intellectual contagion.

Do not misunderstand my criticisms here. I think the academic life can be the most varied and imaginative and interesting life in the world, and I love it. But I am talking about its distortions and about how they narrow it from what it might become. Its great men are so very great and its little men are so little. And it pains me when I see one of those academic men who has deliberately narrowed himself to an intellectual pinpoint and has cut off all that life might be. Emerson must have been thinking of such men when he said: "The state of society is one in which the members have suffered amputation from the trunk, and strut about so many walking monsters—a good finger, a neck, a stomach, an elbow, but never a man."

The academic world is perhaps no worse in this respect than the world of government or the world of business, but it is sad all the same. The teacher is the one man who most needs to know what it is to be a complete man with wholeness and diversity and humor. When his vision is distorted, the vision of a whole generation may be warped.

I think it is time to say loudly and clearly that the interval of higher education should be an interval of learning to live like cultured human beings instead of like monks and academics. Instead of overload and punishment let us have excitement and leadership. Along with excellence let us enjoy diversity. Let us try to find ways in which students can be given the money and leisure they ought to have as valuable apprentices in an affluent society. Let us bring up a generation of young adults full of the delight of living, interested in many things, and knowing not only how to be intellectual but how to be full and creative men and women.

The Second Educational Revolution

I think that this goal I have suggested, of trying to make the college years more humane, more cultured, and more diverse, is just a part of a new educational revolution that will totally change the structure of our schools in the next 20 years. This revolution may be even more thorough-going than the revolution that was made by John Dewey and the other reformers 70 years ago, when they swept out the obsolete and stuffy classical education of the 19th century and redefined the goals of education as education for society and education for living.

Today our education has indeed become an excellent education for our society, so far as its professional content is concerned, but it is still obsolete and clumsy in its teaching methods. Since World War II, a revolution has occurred in information and communi-

cation and in our knowledge of the biology and psychology of the brain and the psychology of learning. It is beginning to be urgent for us to adapt our educational system to take account of these advances. Mass education up until now has been hard and punitive, with more of the stick than of the carrot. It has been hardest and most punitive in the colleges, where many departments and schools are actually proud to have standards so strict that they flunk one-third of their freshmen.

But it is now possible to move away from this traditional pattern. It has become clear that the psychology of positive reinforcement, of encouraged curiosity and reward, works much better than the psychology of negative reinforcement, as great teachers have always known. It is time to try out on a large scale the new discoveries and methods of this new educational psychology, discoveries such as the remarkable effect of early enrichment at ages 1 to 4, and methods such as use of the new phonetic alphabets and the programmed learning and teaching machines and programmed texts that promise to make spelling and geography and physics and anatomy and many other subjects easier and more quickly mastered.

The new ideas have already made a revolution across the nation in the teaching of high-school science courses, and efforts are well under way to create science programs with the same exciting immediacy all the way down to the kindergarten level. In fact it now appears that the whole difficulty with many subjects is that we have been teaching them too late. A 7-year-old can learn reading and writing more easily than an 18-year-old can, and we are now finding that he may also learn about sets and binary arithmetic and rates-of-change and the difference between mass and weight more easily than many college sophomores.

The difficulty today is that these remarkable new methods have not yet been drawn together into a unified educational approach. We have a better engine, a better transmission, and a better steering mechanism, but they have not yet been fitted together to make a complete car. It seems very likely that, when they are all put together, these new developments in education will reinforce each other and will make possible further gains that would not come from any one alone. Pre-school reading and writing would make room for beginning science in the early grades. Binary arithmetic in the second grade may make a child ready and eager for number theory and computer programming in the sixth. Rates-of-change at age 7 would permit introduction to economics at 13.

What is evidently needed now is to get out of the rut of our standard educational structure and to set up complete new kinds of pilot schools to try out this new personal and concrete and manipulative education in an integrated program all the way from age 1 to age 21 and beyond. We need to try schools of several different kinds, in different types of communities, in slum areas and rich suburbs, in company towns and scientific laboratory communities, to find out which kind of program under different circumstances

produces the most alert and creative citizens. If we can find some educational leaders who will take the initiative in establishing private schools of this sort, or who can persuade some forward-looking school boards to try them out, this may be the most exciting educational adventure of the next decade.

I think that, if we put together all the speed-ups and simplifications that these new methods make possible, the children in such schools would no longer be overworked. The subjects we now teach them might be mastered in a much shorter school day, perhaps no more than 3 or 4 hours. There would be less boredom and resistance in school and more time for creative leisure outside. Some parents may shudder at this, because they do not want the children home half the day. But, with the new trends of productivity and automation in our adult life, perhaps creative leisure is one of the things we need to teach children earliest. And, if we let the adult's leisure enrich the children's leisure, homework might even become home play. The interaction between the generations might make for better relations than we have had for years. In fact the children, with their shorter hours, going home from school may soon meet the adults, with their new leisure, going back, hoping to learn in a more voluntary and serious way the subjects they missed in all their years of report-card education.

All this would change our stereotyped pattern of education in a remarkable way. The intense program of work now imposed across a few years in the late teens—where we have to study all day and all night because the earlier grades have taught us so little—might be replaced by an easier longitudinal pattern that would start with easy and fast learning methods at age 1 or 2 and would then go on all our lives for 2 or 3 or 4 hours a day. The children and the college students and the leisured adults might acquire a new attitude toward education. Formal teaching might blend inseparably into more individual and creative leisure-time activities, such as building boats together or learning music or ballet or skiing—or studying embryos and catching striped bass before dawn. Education would be by contagion and long discussion, and the generations might learn to talk to each other again.

A lifetime ago we made the transformation to education for living. It is time now to make the transformation to education for wholeness, for delight, and for diversity.

7.4 *Brainstorming and Orneriness as Facilitators of Creativity*

HENRY CLAY LINDGREN AND FREDRICA LINDGREN

Reprinted from *Psychological Reports,* 1965, **16,** 577–583, with permission of the Southern Universities Press, Inc. The authors have collaborated as a husband-and-wife team on a number of psychological studies.

What are some of the innovative ways in which students can be encouraged to become more creative? Research in social psychology shows that small groups have certain characteristics that can facilitate some kinds of behavior associated with creativity. For one thing, interaction in such groups has a generally arousing, stimulating, and energizing effect. For another, it encourages risk-taking. These two characteristics help explain the major finding of this experiment: brainstorming in small groups tends to raise the creativity level of students.

Also of interest in this study is the finding that more creative subjects tended to express more negative—"ornery"—traits than do those who are less creative. This outcome is consistent with those of other studies that show teachers and fellow students as being less accepting of highly creative students than less creative ones. Everyone, it seems, approves of creativity, but not of creative people.

An interesting minor trend that has appeared in the social and behavioral sciences during the past few years is the increasing amount of attention paid to creativity. The extent to which this trend has affected the activities of psychologists is indicated by the sharp increase in the number of papers dealing with "creativeness" and "creativity" reported in *Psychological Abstracts* after 1957. Between 1951 and 1957 the number of publications listed under these headings averaged about 10 per annum. This number more than doubled in 1958, and in 1959 it jumped to well over 20. Since 1960, *Psychological Abstracts* has been averaging about 40 papers a year in these two categories and the number appears to be increasing steadily. This upsurge in interest may be a by-product of the "space race" with the Russians, or it may be the result of shorter working hours, greater affluence, and more leisure time. Or perhaps there is, as some say, a profound and wide-spread feeling of dissatisfaction with the practical, everyday aspects of life, and a reaching out for something that is less well defined in quantitative, economic terms. Whatever the reason, a great deal of concern has been expressed by leaders in the fields of education, science, business, and government about the need to identify and encourage creativity.

One of the earlier proposals for the stimulation of creativity was made by the advertising executive, Alex F. Osborn, who in 1939 introduced a technique he called "brainstorming," essentially, a group-participation method of problem-solving, in which participants considering a certain problem or question are encouraged to bring up any "crazy ideas" that come to mind, without fear of criticism. Osborn (1957) maintains that "the average person can think up twice as many ideas when working with a group than when working alone." Although brainstorming has attracted considerable attention, especially on the part of people in business and industry, Osborn's claim has not been supported by research. A study by Dunnette, Campbell, and Jaastad (1963), for example, confirmed earlier findings by Taylor, Berry, and Block (1957), that people working alone were more creative than when they were brainstorming with a group. However, Dunnette, *et al.* made an additional finding that suggests that brainstorming may nevertheless have some merit. Those of their *S*s who worked alone *after* group brainstorming did better in this phase of the experiment than those who worked alone *before* brainstorming, thus suggesting that brainstorming in a group might have some kind of facilitative effect on individual performance. One of the objectives of the present study was that of investigating this effect further.

Although the research of Dunnette, *et al.* suggests that group interaction may have some kind of facilitative effect on creativity, creative people themselves often express sentiments that are at odds with this finding. Many even take the position that participation in groups tends to inhibit, rather than facilitate, the creative process. This point of view is indirectly supported by research conducted at the Institute of Personality Assessment and Research (IPAR) at the University of California in Berkeley. Reports emanating from IPAR generally show more crea-

tive people are less dependent on others for behavioral cues than less creative people (Crutchfield, 1955). It is quite possible that this tendency toward independence on the part of more creative people may lead to expressions of antipathy or hostility toward anything that would lead to or be consistent with participation in groups.

One bit of research from IPAR that bears tangently on this behavioral pattern is a study reported by Barron (1963) who compared the self-perceptions of 18 male graduate students who preferred "asymmetrical" paintings (such as those by Picasso, Modigliani, and Toulouse-Lautrec) with those of 18 who preferred "symmetrical" paintings (such as those by Botticelli, Corot, and Gainsborough) in terms of the items checked on Gough's Adjective Check List. The "asymmetrical" group seemed to have self-images that could be characterized as "ornery" or "cantankerous," for they tended to pick adjectives describing themselves as bitter, irritable, pessimistic, dissatisfied, sarcastic, demanding, temperamental, and the like, in marked contrast to the "symmetrical" group, who tended to describe themselves as gentle, patient, peaceable, timid, modest, responsible, and the like. Although Barron did not say whether creative people are inclined to perceive themselves as did the "asymmetrical" group, such a mode of perception would be consistent with their reported tendency to display independence in attitudes and in their social behavior. Furthermore, previous IPAR research by Hall (1958) shows that, like *S*s in Barron's "asymmetrical" group, creative people tend to prefer art forms that are characterized by complexity and imbalance.

The availability of these lists of "ornery" and "non-ornery" self-descriptive adjectives suggested that it might be worthwhile to determine whether any relationship existed between orneriness and creativity, to fill in the gap, as it were, in the research reported by Barron and Hall. In other words, is the orneriness reported by persons preferring asymmetrical art also characteristic of creative people (who also prefer asymmetrical art)? A positive finding might, however, run the risk of contradicting the results of Dunnette, *et al.*, namely, that creativity can be stimulated by interacting with others in a group setting, since orneriness and asociability might interfere with an individual's ability to benefit from group interaction.

In order to shed some light on these questions, the present study proposed to test the following somewhat incompatible hypotheses. (1) Creativity, as indicated in productivity in individual brainstorming sessions, will be increased following group brainstorming sessions. (2) Creativity, as indicated by productivity in individual brainstorming sessions, will be positively correlated with tendencies to view oneself in ways characteristic of persons preferring asymmetrical art—ways which may be described as "asocial" and "ornery."

Method

*S*s participating in the study were 81 men and 86 women enrolled in 8 undergraduate classes in psychology at San

Francisco State College. All *S*s first took the Asymmetrical Preference Test (APT), a 30-item, forced-choice, self-rating test designed for this study and composed of adjectives drawn from the study by Barron (1963) cited above.

*S*s were then given instructions for the brainstorming series. They were directed to write as many captions or titles or "whatever comes to mind" for the picture as they were able. Stimulus material for brainstorming consisted of three "free-form" drawings or cartoons depicting, respectively, a man and a woman, two infants, and an elephant and an ostrich. In each cartoon, the figures faced each other in such a way that it would be easy to imagine them in conversation. The drawings were presented to *S*s in random order during three stages or phases, as follows.

In Phase I (Individual Brainstorming) each *S* was handed a drawing and a sheet of lined paper and given 6 min. to write as many responses as possible, according to the instructions already given.

In Phase II (Group Brainstorming) *S*s were divided into groups of 4 or 5. Each group was given new drawings and a sheet of paper and was told to follow the initial instructions, except that they were to function as a group and were given 10 min. in which to produce as many responses as possible.

In Phase III (Individual Brainstorming) *S*s returned to their original seats, were given a third drawing, and told to follow the same directions as in Phase I. The administration of the APT and the three brainstorming sessions took place within one class period.

The APT was scored by a key corresponding to the self-descriptive adjectives chosen by those *S*s in Barron's study who preferred asymmetrical art. The split-half reliability of this test proved to be .89.

The papers completed by *S*s during Phases I and III were scored according to the number of responses produced. The over-all level of creativity expressed by each of 61 men and 49 women (students in 6 of the 8 classes), was also rated by three judges, operating independently, who classified each set of responses on a 7-point scale, using a method that forced judgments into an approximately normal distribution. The inter-judge reliability was .58. Each individual's level-of-creativity score for each of the two phases (I and III) consisted of the sum of the three judges' ratings.

Results

Number of responses

Table 1 shows that women's responses increased in number during Phase III, but men's did not. A further analysis of the data showed that there was little differences between the sexes in the percentages of those who increased their responses in Phase III (61% for men vs 69% for women), but the mean gain for men was minimized by the fact that twice as many men (32%) as women (16%) *reduced* the number of their responses in Phase III. The difference between the proportions of men and women reducing the number of responses was significant at the .01 level, using *t* tests.

TABLE 1. Means and Standard Deviations for Asymmetrical Preference Test (APT) Scores, Number of Responses in Individual Brainstorming, and Level of Creativity Achieved in Individual Brainstorming

Variable	*N*	*M*	*SD*	*p*
APT: Men	87	10.86	5.4	NS [a]
Women	86	9.66	5.42	
No. of responses: men	87			
Phase I		8.1	4.4	
Phase III		9.04	4.48	NS
No. of responses: women	86			
Phase I		7.6	4.28	
Phase III		9.86	4.42	.001
Creativity level: men	61			
Phase I		11.84	3.91	
Phase III		13.22	3.77	.05
Creativity level: women	49			
Phase I		11.83	4.31	
Phase III		13.48	3.97	.05

[a] By *t* test.

Quality of responses

The level of creativity expressed by number of responses for both sexes increased significantly in Phase III, as Table 11 also shows. Differences between the increases reported for each sex were not significant.

Relationship of APT to creativity

Although there was no significant relationship between APT scores and number of responses for Phases I and III of the experiment, the relationship between APT scores and level-of-creativity scores was significant, as the data presented in Table 2 indicate. There was no significant relationship between APT scores and *increases* in level of creativity. Further analysis showed, incidentally, no significant relationship between the number of responses and the level of creativity.

Discussion

The results generally support the first hypothesis and are consistent with the findings of Dunnette, *et al.,* cited above. Creative quality of responses after group interaction was higher for both sexes, although the number of responses was higher only for women. The fact that twice the proportion of men decreased their responses may be worthy of further investigation, although it does not appear to have had any effect on the quality of their responses as a group.

TABLE 2. Pearson Product-Moment Correlations with Asymmetrical Preference Test (APT) Scores

Measures	Men N = 61	Women N = 49
No. of responses		
Phase I	.06	−.10
Phase III	.10	−.04
Level-of-creativity scores		
Phase I	.24*	.30*
Phase III	.36**	.27*
Increases in level of creativity	.10	−.12

*$p < .05$.
**$p < .01$.

The reasons why group brainstorming subsequently facilitates individual creativity are probably complex. It may be that some kind of general principle of social stimulation operates to produce this effect, or it may be simply that group brainstorming helps to raise a participant's interest level in the task at hand. Still another possibility is that group brainstorming serves as a kind of learning situation in which the individual gets a better idea of the nature of the task and learns how to produce better ideas.* What ever the reason for this facilitation, it appears to have possibilities for further research and for experimental application as an educational or training technique.

The results also support the second hypothesis. The positive product-moment correlation between the level of creativity of responses and APT scores suggests that people at the more creative end of the continuum tend to have self-perceptions resembling those of Barron's students who preferred asymmetrical art, i.e., self-percepts characterized by asocial feellings and orneriness. This finding helps to fill in the triad of positive relationships between creativeness, preferences for asymmetrical art, and asocial or ornery self-regarding attitudes. Evidently the three are likely to go together.

It is interesting to note that the asocial character of such self-percepts proved to be no barrier to *Ss*' ability to benefit from participation in group brainstorming. Although a literal reading of their self-descriptions might lead one to assume that group participation would reduce their level of creativity,

*It is possible, of course, that the observed improvement in level of creativity might have been due to practice. The inclusion of a control group completing three trials without brainstorming would have improved the experimental design. This omission was noted by the authors subsequent to completing the work on which the present paper was based. By that time they had gone abroad for an extended stay and were unable to locate a culturally equivalent group of *Ss*. (A control group actually was included in a later experiment, similar in design, which produced results consistent with the present study [Lindgren and Lindgren, 1965].).

the opposite actually occurred. Such a finding leads one to wonder whether statements that creative people often make about turning their backs on society can be taken at face value. It may well be that social participation contributes more to their potential for creativity than they are aware of or are willing to admit.

Further research regarding the relationship between group interaction and creativity would, of course, be needed before any general conclusions are formulated, but if the findings of such research turn out to be consistent with those reported in the study, it seems that those individuals who are interested in fostering and encouraging the development of creativity should consider using techniques like group brainstorming as facilitators and stimulators of the creative process.

The fact that creativity appears to be associated with certain asocial and ornery attitude patterns poses a different kind of problem, because it is not clear whether such attitudes contribute to creativity or whether they are the result of being creative. One point to keep in mind, of course, is that the magnitude of the correlation between this attitude dimension and creativity, though statistically significant, is not high. Another point worth noting is that it is possible for a person to have ornery and asocial attitudes without being creative. One possible interpretation of this latter, rather unoriginal observation is that the orneriness of more creative persons (or the motivational pattern underlying this orneriness) does not interfere with their being stimulated by group interaction, whereas with ornery people who are less creative, it may interfere.

Summary

When 173 *S*s were assigned the task of writing captions for cartoons in a three-stage experiment in which they brainstormed alone, then in small groups, and finally alone, the level of creativity of responses was higher in the third phase, as contrasted with the first phase. The number of responses produced in the third phase was significantly higher for women, but not for men. Level of creativity, but not the number of responses, was positively correlated with scores made on a forced-choice questionnaire consisting of pairs of self-evaluative adjectives scored in terms of choices made by persons preferring asymmetrical to symmetrical art. Such choices were generally consistent with what may be described as asocial or ornery attitudes.

References

Barron, F. *Creativity and Psychological Health.* Princeton: Van Nostrand, 1963.

Crutchfield, R. S. Conformity and character. *Amer. Psychologist,* 1955, **10**, 191–198.

Dunnette, M. D., Campbell, J., and Jaastad, K. The effect of group participation on brainstorming effectiveness for two industrial samples. *J. Appl. Psychol.,* 1963, **47**, 30–37.

Hall, W. B. The development of a technique for assessing esthetic predispositions and its application to a sample of professional research scientists. Unpublished paper read at Western Psychological Association, April, 1958.

Lindgren, H. C. Brainstorming and the facilitation of creativity expressed in drawing. *Percept. Mot. Skills,* 1967, **24**, 350.

Lindgren, H. C. Creativity expressed in drawing and its relationship to orneriness. *Percept. Mot. Skills,* 1968, **26**, 490.

Lindgren, H. C., and Lindgren, F. Creativity, brainstorming, and orneriness: a cross-cultural study. *J. Soc. Psychol.,* 1965, **67**, 23–30.

Osborn, A. F. *Applied Imagination.* (Rev. ed.) New York: Scribner, 1957.

Taylor, D. W., Berry, P. C., and Block, C. H. *Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?* Yale Univer., Department of Psychology, Office of Naval Research, Tech. Rep. No. 1, 1957.

Addendum

Subsequent research with brainstorming techniques has produced the following results: 1. A replication of this study with Middle East university students, using a control group, produced similar findings. 2. Brainstorming appears to facilitate creativity in drawing and sketching in women college students (but not in men). 3. Orneriness appears to be associated with creativity in drawing and sketching for both male and female college students. 4. Brainstorming appears to facilitate creativity in bright sixth graders (unpublished research).

7.5 *Training Children to be Creative May Have Little Effect on Original Classroom Performance*

FRANK E. WILLIAMS

Reprinted from the *California Journal of Educational Research,* 1966, **17**, 73–79, with permission of the author and the *California Journal of Educational Research.* Frank E. Williams is a member of the education faculty of the Portland State University in Oregon.

The research reported by Williams shows that the more creative student is likely to be a good all-around student—that is, he has a superior IQ, gets better grades, receives high ratings on citizenship, and possesses a great deal of information. It is interesting that although teachers were giving good grades and citizenship ratings to these students, they were unable to name them when asked to identify the most creative students in their classes. Creativity is a very elusive trait: we may not recognize it as such when we see it, but when we are asked to evaluate the performance of members of a group, we tend to give the highest ratings to individuals who have shown more creativity by objective measures.

Of major interest in Williams' research, however, was his inability to raise the creativity level of students by his experimental techniques. As he says, ". . . practice and reinforcement had no effect on original performance of the tasks used in the study. . . ." This does not mean that creativity cannot be stimulated or facilitated, but rather that focusing on the responses of the individual may not be a very good method. On the other hand, setting learning tasks that require divergent thinking and providing a classroom environment that is stimulating and that invites risk-taking may work. Such a method was described in the preceding paper by Lindgren and Lindgren. Even a ten-minute brainstorming session was sufficient to raise the creativity level of a majority of the students who participated in their research study.

Creative performance as indicated by the production of original and fluent behavior in school classroom situations may come only as a consequence of teaching subject matter content. Original thinking is an outcome of breadth and depth of knowledge, and the ability for its production depends on the establishment of associations between items of subject matter content. It is suspected that knowledge of subject matter is the key factor for creative behavior, and the important question for classroom teaching becomes that of how to structure knowledge so that new associations among stored information might be developed.

There remains the possibility that practice and training through some external control by the teacher and/or classroom environment may somehow permit the student to learn how to make new associations out of subject matter content. Some evidence indicates that the inhibition of common associations in the classroom as an externally controllable condition might facilitate the production of more remote or original associations. For example, if the teacher simply would not allow the pat answer given in the book to be uttered in the classroom the children might come up with different answers quite unlike the one in the book. By refusing to accept the one and only answer given in the book, grade school teachers may cultivate creative thinking in their classes.

Recent research gives support to the stated position that a crucial element in the production of creative responses is knowledge. It appears that it is not a matter of "how much" intelligence an individual has, but how he learns to use what he possesses. Creativity may depend on the ability to make new or different associations, and associations depend on breadth of knowledge.

Studies with school students show that those who score high on tests of creativity are likely to have a higher than average amount of stored information, that is known as knowledge, and are also able to do something with this knowledge other than merely receive and store it for recall when conditions necessitate its use.

Sufficient data are now available which show that the child who scores high on tests of creativity is likely to have higher than average subject matter grades, is likely to be better informed, displays better citizenship, and knows how to study better than the less creative child. Further evidence refutes the sometimes quoted statement that teachers do not like the original, creative pupil. Instead, the evidence is consistent with the position that the original child, who inevitably brings so many ideas of interest to the class, is generally favored by the teacher in terms of grades. Teachers, like children, enjoy elements that bring novelty and freshness into the school classroom.

However, it has been found that teachers are not able to identify or select their most creative students in spite of the fact that these students are given the highest academic grades as well as teacher grades for citizenship and study habits. This evidence gives no support to the previous statements which have been made by some educators to the effect that teachers generally discriminate against the creative child.

These and other findings are being reported upon from the recent extensive research study conducted in public school classrooms on the nature and trainability of creative thinking among sixth grade children.

The study, sponsored by the U.S. Office of Education, sought to determine the learning effects for creative performance on well known tasks calling for original thinking; and the relationship of originality and fluency scores on the tasks to grades, intelligence, standardized achievement tests, and teacher selection of creative students in the classroom. Nine sixth-grade teachers and classes involving 294 students were tested before and after intervening training conducted in their classrooms over a five-week period. The Torrance Ask-and-Guess Creativity Test and other creativity measures of divergent thinking were initiated by the classroom teachers before and after training. Variations of the Guilford Unusual Uses Test were given for fourteen consecutive days during regular class periods by an experimenter while conducting a short training program designed to evoke original ideas for the creative use of familiar objects. Originality and fluency scores on all measures were obtained and correlations between these and other pupil performance measures such as IQ and grades were computed. Results are as follows.

Creativity Versus Knowledge

There was a considerable degree of reliability of originality scores across all classes and test days. Even though practice and reinforcement had no effect on original performance of the tasks used in the study, some children were consistently original. The trait designated "originality" as measured in the study appears as a relatively stable trait in a given child.

It may be that one does not train directly for original responses but that other things lead to creative behavior. The results of the study indicate that things other than external conditions such as freedom and opportunity for creating original responses, cues and instruction to be imaginative, and verbal rewards for the most unusual uses are necessary. The effect of external environmental conditions may facilitate creative performance only after the proper internal conditions are attained. It is very clear from a close examination of the responses produced on the tasks that the production of original responses is highly related to the amount of knowledge possessed by a student. The amount of stored information of the person who is attempting to make an original response seems to set the limit on such behavior. Internal conditions which facilitate originality appear to depend on a storehouse of knowledge within the person who is attempting to make a really unusual response. For example, a student who was asked to find an original use for a pencil suggested that the carbon might be used in building a miniature atomic furnace was identified as a highly knowledgeable youngster.

The arrangement and manipulation of external classroom conditions optimum for creative behavior may have its effect only after sufficient knowledge

has been imparted and stored. Practice in evoking original responses is hardly likely to increase the amount of stored knowledge. It appears that training may have indirect effects but creative behavior may come only as a consequence of direct acquisition and storage of a depth and breadth of knowledge. The study does not support the position found in some educational circles that massive generalized transfer of training can be achieved in areas of creativity and original thinking even with the careful control that the study offered.

Creativity and Intelligence

With regard to the relationship between creativity and intelligence found in this study, an extensive review of recent research literature on creativity reveals that the following relationships fit other data quite well. The correlations between measures of originality and fluency with intelligence were all positive and highly significant (ranging between .27 and .45).

In spite of the popular claims made concerning little or no relationship between creativity and intelligence, the findings on this study, for example, match almost identically the findings of Bish (1963) who obtained positive and significant correlations between verbal creativity and intelligence also for sixth grade students varying from .29 to .36 by the California Test of Mental Maturity, the same test as that used for measuring the intelligence of experimental students in the present study. Yamamoto reports (1961) that all of the available research results support the conclusion that correlations between measures of creativity and intelligence range between .20 and .40 in a general unselected population of students. The range of IQ in this study of sixth graders was from 84 to 146.

Torrance (1961) reports several correlations of intelligence between his creativity tests and various forms of intelligence measures. For example, for elementary school pupils the correlation between the Otis Quick-Scoring IQ Test and a creativity score was .32; for sixth grade students using the Kuhlman-Anderson Test it was .26; for fifth and sixth grade students using the California Test of Mental Maturity it was .24; and for high school students using the Lorge-Thorndike he reports .27. The correlations between creativity and intelligence found in this study and a multitude of others give further support to the position stated previously that a crucial element in the production of original responses is knowledge.

Creativity and Grades

Of particular interest are the correlations found in this study between originality scores, subject-matter grades, and grade placement on a standardized achievement test. Originality scores were highly and significantly correlated with a composite grade in the subjects of mathematics, science, and social studies. Individual grades in these three subjects were grouped together to form a composite grade in these subject-matter fundamentals in order to seek their relationship to originality. A likewise composite grade was obtained by grouping individual grades from the

more creative or self-expressive areas of language arts, music and art where a higher relationship with originality was expected. The correlations obtained, however, were almost identical (.375 and .376). When originality scores were correlated with grade placement on the Science Research Associates Standardized Achievement Test, a coefficient of .314 was obtained which was also highly significant.

It is apparent from these data that there was a signifificant relationship between subject matter grades and standardized achievement test scores to tasks calling for creative thinking.

Thus it is that items of information or knowledge become the crux of the problem of being able to form new associations for the making of original responses.

The findings of the study with respect to intelligence and grades suggest that creative students are well above average on those factors which measure knowledge. Their minds have an unusual capacity to receive and store information. The intelligent student is more discerning, is more alert, is more fluent in scanning stored information and producing associations which meet some problem solving criterion. Such a student will generally have more general information at his command. Items of knowledge which he does possess may more readily enter into new combinations among themselves, and the number of possible combinations is increased by greater information and greater fluency of combination. The more combinations or associations between elements of knowledge, the more likely it is that some of them will fit the criterion of infrequency of response and be judged creative. Range of subject matter information and breadth of knowledge seem to be the crux for creative performance.

Techniques for Creative Thinking

Creative behavior may come about as a consequence of mastering the basic facts of a subject, gaining all the information and knowledge first, as a result of direct training. Acquisitions of knowledge first may have a far more powerful effect on originality than all of the creative training procedures which have been devised and experimented with to date. It may be the process of assimilation in which the student develops a whole repertoire of subject matter concepts which give meaning and organization to bits of knowledge.

Original or creative ideas, however, cannot necessarily be guaranteed to follow after assimilation and mastery of information. Assimilation alone is insufficient. Whenever the student encounters information which does not fit his existing concepts, an alteration or extension of his prior stored concepts takes place which is the process of accommodation. The student has to accommodate his conceptual structure to encompass new information which does not fit his current scheme. The processes of exploring new associations which are the necessary prerequisite to making discoveries must be granted students who have abilities for autonomous or creative thinking. The two mental operations of assimilation and

accommodation may contribute suggestions for the development of creative behavior.

Creativity and the Curriculum

The creative process probably requires teaching strategies which guide the development of cognitive skills as well as the many aspects of divergent thinking. Subject matter content must be presented at differing conceptual levels which involve aspiring toward novelty, showing fluidity in associations, showing flexibility, and probing new dimensions of subject matter knowledge.

The creative process also requires challenge and opportunity in the classroom where there is acceptance that this behavior is desired and approved, as well as ample experiences provided for its release. Making new associations after having gained sufficient knowledge more than likely will depend on a multitude of opportunities provided by the teacher for allowing novel associations to become a part of daily classroom behavior. To help students in the ability of forming new associations certain opportunities can be provided such as allowing them to apply their knowledge in new situations, raising questions which focus a query but do not prescribe ways of getting at them, to compare and contrast, and to find differences as well as similarities. It seems, then, that creative responses ter grades, is likely to be better in-edgeable events but do not necessarily have to follow unless conditions are suitable. There is evidence also that even in a creative environment with the proper kinds of opportunities there are great differences among students for putting ideas and factual knowledge together in original associations.

Teachers' Selection of Creative Students

When those teachers in the study were asked to identify and select their five most creative students it was found that they were unable to do so. Whatever the classroom teachers' criterion for an original student may be, it is apparently not the same as that which is being measured by creative tests used in the study. In essence, none of the teachers were selecting the same students on the basis of originality that the test measures did. The data indicate that there is little agreement between a teacher's judgment of original and fluent students and what the particular creativity tests used in the study measure.

One would wonder if ability to select creative children might become possible if teachers were exposed to some in-service training on what is currently known about the creative child and process, or if the teachers themselves were more disposed toward creativity ability. The problem may be one of properly orienting teachers to what it is they are looking for. This is now being tried out on a new group of teachers by means of in-service training seminars on the subject of creativity.

The task of teachers may be that of attempts to structure or reorganize subject matter to include all of the fluencies, flexibilities, elaborations, and

open-ended interpretations of knowledge that comprise the creative process. For example, teaching subject matter creatively might merely involve using various fluency factors of several kinds, such as flow of ideas, flow of associations, and flow of expressions for stating ideas or solutions. The implications of the study make a plea for teachers to take a new look at knowledge when attempting to evoke creative performance rather than to just add more knowledge to what is already there. Knowledge learned yesterday can help form today's new associations under situations calling for creative performance, and creative performance today tests yesterday's amount of learned knowledge.

References

Bish, Gertrude G. A study of the relationships of intelligence, achievement, creativity, anxiety, and confidence among intermediate grade pupils in a suburban area elementary school. Unpublished doctoral dissertation, George Washington University, 1963.

Torrance, E. P. Factors affecting the development of the creative thinking abilities. *Merrill-Palmer Quarterly,* 1961, **7**, 171–180.

Williams, F. E. Practice and reinforcement as training factors in creative performance. Unpublished doctoral dissertation, University of Utah, 1965.

Yamamoto, K. Creativity and intellect: review of current research and projection. Paper read at Minnesota Psychological Association Convention, April, 1961.

Van Bucher

PART 8

Providing for Individual Differences

8.1 *Effects of Accelerating Bright, Older Elementary Pupils —a Second Follow-up*

HERBERT J. KLAUSMEIER, WILLIAM L. GOODWIN, AND TECKLA RONDA

Reprinted from the *Journal of Educational Psychology,* 1968, **59**, 53–58, with permission of the authors and the American Psychological Association, Inc. Herbert J. Klausmeier is director of the Research and Development Center for Learning and professor of educational psychology at the University of Wisconsin. William J. Goodwin is at Harvard University, and Tekla Ronda is with the public schools of Racine, Wisconsin.

The problem of what to do with bright students who race through their assignments and perform far above the level of the average student is one that has plagued public school teachers from the very beginning. Giving longer assignments is not the answer, nor do "enrichment programs" and ability grouping seem to work. Teachers and administrators are understandably reluctant to accelerate (double-promote) bright students, yet research shows that this is the only treatment that produces consistently positive results. The findings reported in the paper by Klausmeier, Goodwin, and Ronda, the latest of a series of observations of two groups of Racine school children who "skipped" the third or fourth grade after having completed a five-week summer "cram session." Like other studies of educational acceleration, the report from Racine shows that nothing is lost and much is gained when bright students "skip grades."

The ideal of public education is to encourage each child to learn as well and as fast as he can, commensurate with optimal personality development. Lack of complete success in achieving this ideal is shown in the widespread attention given to academically talented children in the late 1950s and early 1960s and, more recently, to educationally disadvantaged children. This attention to various groups of children clearly indicates that instruction has not yet become sufficiently individualized to provide well for each child. Therefore, special provisions must be made on a widespread basis for groups of children. One of many possible provisions for the gifted that offers effective utilization of the resources of the school with little change in grouping procedures, organization, and the like, is acceleration whereby the student completes 12 grades of school in less than 12 calendar years.

In 1960, a random half of all the bright, older pupils who met certain specified criteria and who had just completed the second grade in Racine, Wisconsin, were accelerated to the fourth grade after a 5-week summer session. During this session, instruction related to the usual third-grade curriculum was given. Toward the end of the fourth grade,effects of this acceleration appeared to be entirely favorable (Klausmeier & Ripple, 1962). During the summer of 1961 the random half of the bright, older pupils that had not been accelerated were given the opportunity to accelerate from Grade 3 to Grade 5 after participating in a similar 5-week session, but they were not studied further. However, the first group accelerated from second to fourth grade was studied intensively again and was doing well toward the end of the fifth grade (Klausmeier, 1963). The present study considers the longer-range effects of the experience upon both groups of accelerants toward the end of the ninth grade.

Method

Subjects

The 129 *S*s (54 boys and 75 girls) were distributed as shown in Table 1. The abbreviations used in Table 1 and below refer to groups as follows:

Acc 2–4—Accelerated from second to fourth grade in 1960, currently ninth-graders.

Acc 3–5—Accelerated from third to fifth grade in 1961, currently ninth-graders.

9SY—Nonaccelerated pupils of superior ability below the median age of normally progressing ninth-graders.

9SO—Nonaccelerated pupils of superior ability above the median age of normally progressing ninth-graders.

9AY—Nonaccelerated pupils of average ability below the median age of ninth-graders.

9AO—Nonaccelerated pupils of average ability above the median age of ninth-graders.

The students in the Acc 2–4 and Acc 3–5 groups were identified in the spring of 1960 and the others in the fall of 1960. At that time the Acc 2–4 and Acc 3–5 groups each had 16 girls and 10 boys. The same numbers of boys and

TABLE 1. Characteristics and Distribution of Subjects

Characteristic	Group					
	Acc 2–4	Acc 3–5	9SY	9SO	9AY	9AO
Mean IQ September, 1960	123.77	121.58	123.77	124.62	103.77	100.04
Mean deviation IQ May, 1966	130.9	127.6	125.0	129.9	110.1	111.4
Average age on September 1, 1965	13–5	13–5	13–11	14–3	13–11	14–4
Male subjects	9	6	11	9	8	11
Female subjects	13	8	16	13	13	12
Total	22	14	27	22	21	23

girls comprised the other groups for statistical analyses although two alternate girls and boys were also identified in each group. In the present study the alternates were included in order to alleviate somewhat the natural exodus that had occurred during the six years since the study was started. As may be noted in Table 1, the smallest group is Acc 3–5 ($N = 14$). This group is smaller than Acc 2–4 partly because fewer were accelerated from Grades 3 to 5 than from Grades 2 to 4.

Also presented in Table 1 are Kuhlmann-Anderson IQ scores as of September 1960 and the mean IQ scores in deviation form as obtained from the Kuhlmann-Anderson Test, Seventh Edition, Booklet H, Personnel Press, Inc., in May 1966. The differences among the first four groups were not significant at the .05 level in 1960 nor in 1966. It is interesting to observe that the mean IQ of each group increased. One cannot determine whether the increase is solely a test artifact or whether these groups actually increased in IQ above the national standardization sample.

The mean age of each of the six groups when they began ninth grade is included in Table 1. Both Acc 2–4 and Acc 3–5 will be graduated from high school in June with a mean age of 17–2. The 9SY and 9AY groups will graduate at a mean age of 17–8, while the 9SO and 9AO groups will finish with a mean age of 18–0 and 18–1, respectively.

Instruments used and treatment of data

Five types of data were assembled on the pupils near the end of Grade 9. A brief description of each instrument follows.

Educational achievement. The Tests of Academic Progress, Grade 10, Form 2, Houghton Mifflin Company, were administered to all subjects; Grade 10 of the test was given to provide an adequate ceiling for the ablest students. This test yields scores in six areas: social studies, composition, science, reading, mathematics, and literature. Raw scores were converted to standard T scores ($M = 50$, $SD = 10$) by means of the test manual; Grade 10 students were used as the norm group in conversion. Thus both accelerant groups, now in the ninth grade, are reported as having standard score means of 52 on

the reading subtest since this is based on tenth-grade norms.

Ingenuity in problem solving. Form A, Ingenuity, of the Flanagan Aptitude Classification Tests (FACT), Science Research Associates, Inc., was utilized as a measure of ingenious problem solving. Although this test has a large verbal component and a rather restrictive format for gauging "ingeniousness," it nevertheless allows objective scoring. Each item right is scored 1; a maximum score is 25. High school seniors were used to norm the test, with raw scores of 14, 18, 21, and 24, falling at the 50th, 75th, 90th, and 99th percentiles, respectively.

Creative thinking abilities. Four instruments yielding eight scores were used: Alternate Uses, Form A, Expressional Fluency, Form A, Consequences, and Plot Titles, 0–1, Sheridan Supply Company. Each of the four tests was scored for fluency (number of relevant responses). Expressional Fluency and Alternate Uses were also scored for flexibility (i.e., number of relevant categories of response), while Plot Titles and Consequences were evaluated for cleverness of response. These tests were originally developed by Guilford and his associates and, although some of them are only recommended for experimental use, some available reliabilities are reported in accompanying manuals. Depending on the difficulty of the judgments to be made, a training session was held and then either two or three scorers, working independently, scored the tests. The available scores on each pupil were then averaged across scorers. The average interjudge reliability in determining each of the eight scores is reported in Table 2.[1]

Psychomotor abilities. Three tests of psychomotor abilities were devised and administered by Grace Piskula, Consultant in Physical Education for the Unified School District, Racine, Wisconsin: Zig-Zag Run, to measure agility and large muscle coordination (the fewer the seconds required to complete the run, the higher the ability); Wall Pass, to determine eye-hand coordination and speed of reaction (the more hits of a wall with a ball in a 15-second interval, the better the coordination and reaction); and Standing Broad Jumps, to judge leg strength and ability to coordinate body parts (scores reported in inches jumped).

Participation in school activities and special programs. Each student involved in the follow-up was given a questionnaire regarding his activities, both in and out of school. The responses on these were tabulated and the percentage of the group responding yes to each item was computed. The areas of interest to investigators will become apparent in the results section, and include matters such as enrollment in "condensed" courses, honor roll lists, and participation in nonclass activities.

A 6 $\times$ 2 analysis of variance (groups by sex) was run on each of the measures except those under school participation. The latter are reported and discussed as percentages. Where the

[1] Table 2 has been omitted in the interest of brevity. It reported reliabilities between .62 and .85 for "cleverness," .95 for "flexibility," and .94 to 1.00 for "fluency."

TABLE 3. Means, Standard Deviations, and Significance of F *Ratios for 18 Measures*

Measure	Acc 2–4		Acc 3–5		9SY		9SO		9AY		9AO		Significance of *F* Ratios	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Groups	Sex
Tests of Academic Progress														
Social studies	56.6	8.4	57.3	8.2	56.1	6.9	62.0	5.8	47.0	7.9	47.9	6.6	.001	*ns*
Composition	58.3	8.2	54.4	7.7	57.0	9.1	61.4	6.5	46.0	7.3	48.6	6.3	.001	.01 F
Science	53.5	6.6	52.3	7.5	49.9	6.6	56.8	8.3	46.4	7.6	47.9	5.5	.001	.05 *M*
Reading	52.0	8.5	51.5	9.5	50.6	8.8	57.2	7.0	44.0	7.4	46.1	7.7	.001	*ns*
Mathematics	62.1	6.8	60.4	7.4	58.1	7.1	61.9	6.3	47.6	6.2	51.3	7.0	.001	.001 M
Literature	55.2	7.7	54.3	9.0	53.7	8.2	60.1	7.3	44.8	7.3	45.9	6.6	.001	*ns*
Flanagan Ingenuity Test	17.2	3.5	16.2	3.5	15.8	3.2	17.4	3.3	14.0	3.6	15.7	2.6	.01	*ns*
Creative Thinking Battery														
Expressional Fluency														
Fluency	8.9	3.4	8.9	3.4	7.8	2.6	10.3	2.3	6.8	3.8	7.5	4.0	.01	*ns*
Flexibility	8.1	2.8	8.7	3.4	7.6	2.5	9.9	2.5	6.4	3.5	6.4	3.3	.001	.01 F
Alternate Uses														
Fluency	25.8	6.0	22.0	6.4	25.2	8.3	26.4	8.6	23.5	7.4	20.5	7.2	*ns*	*ns*
Flexibility	24.3	4.8	21.0	6.3	23.9	7.4	25.2	7.8	20.4	7.9	19.0	7.1	.05	*ns*
Plot Titles														
Fluency	13.1	5.0	13.0	4.3	13.7	5.6	13.3	4.7	12.7	5.5	11.6	3.2	*ns*	*ns*
Cleverness	2.4	1.0	2.2	1.4	2.0	1.3	3.3	1.6	1.6	1.0	1.0	1.0	.001	*ns*
Consequences														
Fluency	50.7	16.2	55.0	19.9	57.7	23.8	57.0	17.4	49.7	18.7	50.6	13.6	*ns*	*ns*
Cleverness	21.8	7.8	22.6	11.5	21.6	7.2	25.9	8.3	17.7	8.4	14.7	4.9	.001	*ns*
Psychomotor abilities														
Zig-Zag Run	24.2	2.5	24.6	2.8	24.6	2.8	23.0	2.9	24.4	2.7	23.6	3.4	*ns*	.001 F
Wall Pass	10.4	1.9	10.7	1.6	11.4	1.9	12.2	2.6	11.3	1.6	11.2	2.7	.05	.001 M
Standing Broad Jump	67.5	10.6	67.6	10.3	69.4	10.2	74.8	13.2	69.0	7.8	73.2	14.6	.05	.001 M

difference among the six groups was significant at the .05 level or beyond, a Newman-Keuls test was run to ascertain which groups were significantly higher or lower than the two accelerant groups. Differences between sets of the nonaccelerant groups (9SY, 9SO, 9AY, and 9AO) are not presented in the interest of brevity.

Results

The means and standard deviations for total males and females in each group are presented in Table 3, as well as the significance level of the *F* ratios for groups and sex, with an indication of whether males or females were significantly higher. The mean scores on the tests of educational achievement are of interest directly because the test battery and the related norms were for tenth graders. A score of 50 is equivalent to the median score attained by the tenth graders who comprised the national standardization sample. The mean scores for both Acc groups and the 9SO group were above the tenth-grade median. A mean score of 60 is roughly equivalent to a percentile score of 85. Thus, the mean score of 62.09 made by the Acc 2–4 group in mathematics indicates very high achievement.

The differences among the means of the six groups were statistically significant on 14 of the 18 measures. The four nonsignificant differences occurred on the creative thinking battery—ideational fluency scores on Alternate Uses, Plot Titles, and Consequences—and on the Zig-Zag Run. Generally, the 9AY and 9AO groups had the lowest mean scores on the 14 significant measures while 9SO had the highest. Especially to be noted is the fact that on 14 of the 15 tests in the cognitive domain, the 9SY group did not differ significantly from either Acc group, both of which had 1 less year of schooling and were on the average 6 months younger than 9SY. Also, the 9SO group was significantly higher than both Acc groups on only two measures and higher than one Acc group but not the other on four measures. Part of the superiority of the 9SO group may be related to a somewhat different pattern of educational experiences, to be considered more fully later. The comparison of the two Acc groups is also of interest because they had been accelerated at different times in their school careers. No difference between these two groups was significant.

The difference between the sexes, independent of the groups was significant on seven tests. On the educational achievement measures, males were significantly higher in science and mathematics, while the situation was reversed for composition. Girls were significantly higher than boys on the flexibility score of the Expressional Fluency test. Boys performed significantly better than girls on all the psychomotor tests (girls took significantly more seconds to run a specified distance and thus performed less well than boys).

On three measures the Group $\times$ Sex interaction was significant: TAP Reading, TAP Mathematics, and the Standing Broad Jump (all $p < .05$). The girls in the Acc 2–4, 9SY, and 9AO groups scored considerably higher in reading than did boys; however, the boys were somewhat higher than the girls in the

other groups—Acc 3–5, 9AY, and 9SO. No explanation for this interaction can be offered either in terms of the composition of the groups initially or their subsequent education. Both ability levels, superior and average, and both programs, accelerated and normally progressing, are equally involved. The Group X Sex interaction in TAP Mathematics was also significant. Here boys in all groups, except 9SY, had higher mean scores than girls; however, the difference between the means of boys and girls varied markedly among the five groups. The significant Sex X Group interaction for the broad jump is related to the unequal differences between the mean scores of boys and girls in the various groups inasmuch as boys were higher than girls in all groups. The large differences between boys and girls occurred in the 9SO and 9AO groups; the smaller differences occurred in the other four groups comprised of younger children. This difference is explainable in terms of physical development, namely, on measures of strength the difference between boys and girls increases with age. (The difference in running speed also increases with age but was not sufficiently large in this study to produce a significant interaction.)

Contained in Table 4 is a summary of the extent of participation by the various groups in school activities and special programs. The compressed courses were formed by presenting the normal subject matter content for 2 years in a single year. Thus, 2 years of social studies are combined in Grade 6, 2 years of mathematics in Grade 7, 2 years of science in Grade 8, and 2 years of English in Grade 9. A student completing all four courses may be taking the equivalent of tenth-grade classes as a ninth grader. As can be seen in the table, about half of each accelerant group and the 9SY group had enrolled in these courses, while three-fourths of the 9SO group had. Thus, the 9SO group had studied more tenth-grade content. The difference in participation between the accelerant and 9SI groups is particularly evident in social studies and science. Participation by the 9AO and 9AY groups in the compressed courses was very limited.

Related to the question of participation in condensed courses is that of enrollment in special summer school programs for enrichment. The percentages in Table 4 denote that half of the Acc 2–4 group, half of the 9SO group and about one-fifth of all other groups (except 9AO with about one-tenth) had attended at least one summer school session. However, the 9SO group attended about twice as many summer sessions as Acc 2–4 group and four times as many as Acc 3–5. This situation, coupled with that in the preceding paragraph, highlights the greater exposure to enriched content that the 9SO group received.

The table also indicates the percentage of each group attaining the honor roll. The percentage presented here and in the three categories below is actually an average percentage; the participation over the 3 years in junior high school has been averaged to present an annual figure. Thus, on the average, 23% of the Acc 2–4 group attained the honor roll each year, 26%

TABLE 4. Summary of Participation in School Activities and Special Programs in Junior High

Occasion	Group					
	Acc 2–4	Acc 3–5	9SY	9SO	9AY	9AO
Percentage who took compressed courses:						
Social studies	68	29	52	86	0	4
Accelerated math	73	57	48	77	0	9
Accelerated science	41	36	30	82	5	0
Accelerated English	41	50	48	55	10	17
Average	56	43	45	75	4	8
Summer school:						
Percentage who attended one or more summer sessions	50	21	19	55	19	9
Average number who attended per group member	.64	.29	.48	1.23	.19	.09
Honor roll:						
Average percentage who attained annually	23	26	17	32	5	4
Varsity teams:						
Average percentage who participated annually	15	11	24	41	13	48
Intramurals:						
Average percentage who participated annually	70	61	36	48	38	42
Activities:						
Average percentage who participated annually	21	19	22	36	17	14

of the Acc 3–5 group, etc. Although slightly more of the accelerated pupils attained the honor roll than did the 9SYs, the 9SO group placed slightly more pupils on the roll than the accelerants. Attainment of the honor roll by average pupils was negligible.

The male accelerants as a group participated in fewer varsity sports than 9SY, 9SO, and 9AO. However, the accelerants participated in more intramurals than any of the other groups, and their overall participation in sports is quite high as it was for all groups except 9SY and 9AY. Whether less participation by intellectually able boys in sports as members of a varsity team is a strength or weakness of the program is debatable at this time. The relative proportions that complete first the baccalaureate and then graduate school programs is probably more important than is current participation in athletics.

The last tabular entry reflects, again with an average annual percentage, the

extent of participation in nine other activities (student council, school paper, club activities, cheerleading, science fair, talent show and other productions, orchestra, band, and chorus). As can be noted, somewhat higher participation was maintained by the 9SO group with about a third of the group being involved in each activity. All other groups had an average involvement of about one pupil of every five in each activity. From the amount of participation by the Acc 2–4 and Acc 3–5 groups, in comparison with other groups, one may infer normal sociability and social development.

Discussion

At the end of the fourth and fifth grades, the effects of acceleration on the cognitive, psychomotor, and affective development of bright older children were considered generally favorable. At the end of the fifth grade, the then 5SO group was significantly higher than the Acc 2–4 group on only two of eight scores of the Metropolitan Achievement Test, none of ten tests of divergent thinking, and one of six psychomotor tests. At the end of the ninth grade, the 9SO group was significantly higher than the Acc 2–4 group on three of six educational achievement tests, one of eight tests of divergent thinking, and two of three psychomotor tests. The increased superiority of the 9SO group in educational attainment in the ninth grade, compared with the fifth, can be partially accounted for on the basis of their having taken more of the compressed courses for high-achievers during the junior high school years and also having taken more enrichment courses in summer programs. They had more opportunity to learn the subject matter included in the tenth-grade test battery used in this study. Their superiority in the physical measures probably is associated with the greater difference in physical development between the ages 14 and 13 in comparison with ages 10 and 9, the respective nearest ages of the two groups as ninth and fifth graders. In turn, the superiority in physical development was represented by the 9SO boys participating in interscholastic athletics to a greater extent.

The 9SY group is a critical comparison group at the junior high school level for they are 6 months older than the Acc 2–4, have had an additional year of schooling (at considerable additional cost), and are normally enrolled in the same classes with the Acc 2–4 and 9SO groups. The Acc 2–4 group is not significantly different from the 9SY on 17 of the 18 measures; on the mathematics test, the Acc 2–4 was significantly higher than the 9SY group. The mean score for the Acc 2–4 group is actually higher on all six educational achievement tests, on the Flanagan Ingenuity Test, and on six of the eight creative thinking tests. The Acc 2–4 group has slightly lower mean scores on the three psychomotor measures. Further, there is little difference between the Acc 2–4 group and the 9SY in a variety of school activities, including enrollment in the compressed classes and participation in social clubs or activities. The Acc 2–4 par-

ticipate slightly less frequently on varsity teams but much more frequently in intramural activities. The preceding comparisons apply about equally well to the Acc 3–5 group; however, they were not studied until the ninth grade.

Not to be lost in the comparisons are the 9AO and 9AY groups. They too, have had an additional year of schooling and are considerably older than the accelerants. On five of six measures of educational attainment they are significantly lower than both Acc groups and lower than the Acc 2–4 group on the other. On none of the 15 measures in the cognitive domain is either of these groups superior to the Acc 2–4 group and only the 9AO group is superior on one of the three psychomotor measures.

These same groups will again be studied in two years and more final conclusions may be possible then. Based upon all the data collected toward the end of the ninth grade, the effects of acceleration are considered completely desirable. Some bright older children should be accelerated during the elementary school years so that they become the younger high achieving members of their classes rather than remaining the older members throughout their school life. One can predict with confidence that they will continue to be high achievers and to participate in many school activities throughout their high school years.

Summary

Bright older children accelerated in lower elementary grades were compared with nonaccelerants toward the end of 9th grade. Ss were 22 children accelerated from Grade 2–4, 14 children accelerated from Grade 3–5, and 4 nonaccelerant groups: 27 bright younger children, 22 bright older children, 21 average-ability younger children, and 23 average-ability older children. On 6 tests of educational achievement, 9 tests of divergent thinking, and 2 psychomotor tests, both accelerant groups were equal to or higher than the other 4 groups. The nonaccelerated older bright children were higher than at least 1 of the accelerated groups on 4 tests of educational achievement, 2 tests of divergent thinking, and 2 psychomotor tests. The accelerated groups participated in school activities, advanced classes, and varsity athletics, to about the same extent as the older bright nonaccelerants.

References

Klausmeier, H. J. Effects of accelerating bright older elementary pupils: A follow-up. *Journal of Educational Psychology,* 1963, **54**, 165–171.

Klausmeier, H. J., and Ripple, R. E. Effects of accelerating bright older pupils from second to fourth grade. *Journal of Educational Psychology,* 1962, **53**, 93–100.

8.2 *The Ombudsman: Relevant Role Model for the Counselor*

MICHAEL A. CIAVARELLA AND LAWRENCE W. DOOLITTLE

Reprinted from the *School Counselor*, 1970, **17**, 331–336, with permission of the authors and the American Personnel and Guidance Association, Inc. Michael A. Ciavarella and Lawrence W. Doolittle are both members of the faculty of Shippensburg State College, Shippensburg, Pennsylvania.

All students develop problems at some time or other that require special attention, and some problems can seriously interfere with learning progress. For many years, school people felt that such difficulties were regrettable but were not their responsibility, but today there is a general feeling that we should do anything in our power to see that students get the maximum from their educational experience. The roles of school personnel workers—counselors, psychologists, and social workers—were specially created to fill in the gap between the impersonal, administrative demands of mass education and the idiosyncratic needs of individual students. Ciavarella and Doolittle point out that student personnel workers do not need to limit their functions to helping students adjust to the school, but that they may also help the "system" adapt itself to student needs.

In the May 1969 (p. 327) issue of *The School Counselor* the question of counselor relevance was raised with these words:

> *You as a school counselor have a special opportunity, a special responsibility to be relevant, to both your students and your colleagues. How are you involved with students' problems with Selective Service, drugs, sex, problem pregnancies? How do you relate to student unrest and educational revolt, racial problems, the generation gap?*

If, as the publication proposes, "being relevant is really what it's all about today," that is, "doing your own thing but meaningfully," how relevant is the school counselor in the following case which has been presented by Noyes and McAndrew (1968, p. 58–59)? A student is speaking:

> *It's a system, you have to understand that. I guess it's because there are so many kids and they all have to be in school so many hours. Or maybe it's because the people who run schools finally get to the point where they don't like kids and don't want to have too much to do with them. Anyway, it's a system. It's like a machine. One person, a person like me, say, can't beat it.*
>
> *Let me tell you about it. I'm failing math and science, see? My second year algebra teacher told me she took the exact same course I'm failing now when she was a sophomore in college. But that was a hundred years ago, and now they won't even let you in college unless you've already had it. I'm in the eleventh grade, and if I fail these two subjects, I probably won't get into college. So I look around and I see that other kids are passing them who aren't any smarter than I am. So I figure the trouble has to be with me. So I need guidance. right? I decide I should go talk with the guidance counselor. But you see, that's where the system comes in. The guidance counselor has the whole eleventh grade to worry about in my school—642 students. So he keeps this sign on the door:* ***Do Not Enter Without an Appointment.*** *I guess he's in there, talking with some kid who's in trouble or something, and doesn't want to be interrupted. I go down there before school, after school, and during my lunch break. The sign is always there.*
>
> *Finally, I go to the main office and ask how I can get an appointment with the guidance counselor. The girl behind the desk, a student, looks at me like I'm not all there and says that I have to make an appointment with him, personally. I explain that I can't get into his office to make an appointment, and she says that the sign's only up before and after school and at lunch period, and that I can get into his office during study hall, if I get a pass from the main office to leave the study hall. Are you following me?*
>
> *O.K., fine. The only trouble with this is I don't have a study hall. I'm carrying a full load and I'm in some class or other all day. Well, the girl didn't know what to do about this, so she went and asked the old lady who works in the office. The old lady comes over to me and starts in again at the beginning and tells me that appointments with the guidance counselor must be made during the student's study hall, and she*

says that it works out all right because students who don't have study halls don't have them because they are carrying a full load, and the only kids who are allowed to carry a full load are the smart ones who don't need the counselor anyway.

You see what I mean? You just can't beat it. It's kind of funny when I tell it like this, like a comedy of errors or an old Laurel and Hardy comedy on TV. But it's not funny to me because it's my life. *If I don't get into college, I'll probably get drafted and get my head blown off for reasons I don't understand. I don't know what's going to happen to me. How can I keep my sense of humor when I'm going to get ruined by a damned* system?

What good is it for a counselor to have accurate, useful information about the Selective Service, drugs, sex, problem pregnancies, student unrest, racial problems, the generation gap, and career planning if the recipient of the information, the counselee, cannot get to see the counselor? What good are all the counseling techniques in the world if there's a sign on the counselor's door that says Do Not Enter Without an Appointment? How *relevant* is the school counselor in such a school system?

Supposedly, the focus of counseling is on the individual. Yet in today's highly bureaucratized school, the individual is frequently lost in the morass of the organization. The personal touch is gone. The student now becomes a pink card for English, a red card for Math, and a green card for Science. Victimization by the almighty computer is complete, and there is no one to whom the student can turn, not even a counselor.

In the case above, no one in the school wanted to thwart the needs of the student involved. But in his failure to see the busy counselor, the student perceived the counselor as part of an all-encompassing system designed to suppress his individuality. In the student's view, the counselor, like the system, was out to ruin him. When students feel this way, or when they have to appeal to their parents to save them from the system, our schools are weakened.

Because the counselor more than any other person in the school is concerned with the individual—as an individual—it would seem that upon his shoulders would fall the responsibility of assuring that the system does not destroy the individual. It may be that out of all the roles the counselor plays, the one in which he acts as "protector of the student against the system" is probably his most important one.

The Ombudsman

In the Scandinavian countries a similar role has been molded into a position called "The Ombudsman." Bexelius (1968, p. 24) presents a succinct but clear description of the Swedish Ombudsman: "The Ombudsman is an officer of Parliament who investigates complaints from citizens that they have been unfairly dealt with by government departments and who, if he finds that a complaint is justified, seeks a remedy." The resolution of complaints does not always win friends for the Ombuds-

man, as Bainbridge (1965, p. 151) points out:

The Ombudsman cannot be concerned about his popularity. It is no secret that high officials in Sweden—all of them—dislike the ombudsman. They say that he is always interfering in things he doesn't know anything about, and that they could do their jobs better if he would stop meddling, and so on. But all their grumbling doesn't mean a thing. Everybody knows that it is necessary to have an ombudsman.

Jackson (1968, p. 154) has described the teacher as a person who has "a dual allegiance to the preservation of both the institution (the formal organization) and the individuals who inhabit it." This is an ombudsman's role. It is through this kind of role that an institution such as a school becomes strengthened by becoming more sensitive to the needs of its clients.

The Campus Ombudsman

An American adaptation of the ombudsman's role is emerging at the university level. In a recent article by Rowland (1969), the role of the "campus ombudsman" at Michigan State University was described. Here is an ombudsman who, with no power and appointed by the university president, depends upon the good will of the institution and its members to perform his role. Yet, even in this exceedingly weak position, the ombudsman has performed his role so effectively that the concept of campus ombudsman has spread to 20 other colleges and universities. Rowland (1969), p. 37) cites these examples as illustrations of how effective the ombudsman is:

One Michigan State student was shuttled back and forth between two offices six times when he tried to obtain a duplicate copy of a fee receipt he has lost. Finally he unloaded his exasperation on Dr. Rust. The outcome is best described in the student's own words: "He phoned someone who immediately found the extra copy. Dr. Rust told me to go back to a certain girl in the Administration Building, and when I got over there, a duplicate receipt card was waiting. The girl informed me that as far as she knew it was the first one ever issued at MSU."

And in another case

A female graduate student who solicited Dr. Rust's help in gaining release from a financially burdensome contract with the University commented afterward how gratifying it was to learn that even in a large institution "there is at least one person who cares."

Implications for the School

These same kinds of problems exist in the school. The student desperately needs someone in the school he can go to who will listen to his complaint and *act* on it. He needs someone who can humanize the system. In the earlier illustration, for instance, someone needed to remind the administrator, the counselors, and the clerical staff—a hidden bureaucratic force in many

schools—that the forbidding sign on the counselor's door was interfering with the student's immediate, human need for help.

What follows is a discussion of an ombudsman's attributes and functions that might be adapted to the counselor's role. It should serve as a beginning model for the development of ombudistic counselors.

Nobility

One aspect of the ombudsman that appeals to many is the aura of nobility that seems to be associated with the man holding the office (Doolittle, 1969). As Abraham (1968) puts it, "his tone of dignified yet tough accomplishments has won for [his office] and him the very kind of acceptance, confidence, prestige, and acclaim that is obviously indispensable to its successful function."

The counselor must have this same nobility. He must stand with dignity between the student and the school. He cannot hide behind clerical work nor can he retreat into the protective cover of inaction under the guise of Rogerian counseling.

Advocacy

A second adaptable function is *advocacy.* Certainly there should be someone in the school who will advocate making changes. The school counselor is the most logical person to perform this function. He is versed in the psychodynamics of student behavior and has access to vital information that attests to the need for change. By advocating early enough he insures that today's sporadic student irritations about the school and its personnel do not become tomorrow's major grievances.

One of the greatest tragedies in the school is the student who has no one —not even a parent—to advocate for him. Such a student is doomed to an existence in school characterized by disinterest, nonparticipation, and failure. The counselor, having made a commitment to help *all* students in the school, cannot let this happen.

Acting

A school counselor, like the ombudsman, must *act* if he is to be relevant. He helps himself very little and the student even less if after a lengthy discussion of the problem he tells the student who desperately needs help, "My, you have a problem!" This is why the student came to see the counselor in the first place. He needs more than a reaffirmation of what he's felt all along. Like the campus ombudsman, the school counselor must act. And it takes more than sympathetic listening to convince the student that the counselor really cares. To prove his sincere desire to help, the counselor may have to recommend, to refer, to evaluate, to point out, to telephone, to investigate, to intervene, to stimulate, to encourage, to promote, etc. Actions *are* louder than words. Action is *doing*—and doing convinces.

Action should be translated into the counselor's making himself available for help rather than succumbing to the

bureaucratic practice of putting up an "Out to Lunch" sign.

Criticizing

The final ingredient so necessary to the role of an ombudistic counselor is criticizing. Constructive criticism improves a system; it does not weaken it. Without criticism the bureaucratic system fails to see where it is failing to serve *all* its clients (Doolittle, 1969). For example, in *Up the Down Staircase* the administrative assistant stressed the importance of "using the library" while at the same time establishing roadblocks that prevented students from going there. It is much like a school that prides itself on having counseling services, and then, through an administrative convenience ("Do Not Enter Without an Appointment") prevents students from using the services.

It is this type of criticism which, if constructive, benefits the school. No counselor committed to helping students can idly sit by and support a school which, on the one hand, expounds the virtues of attaining certain goals while on the other hand makes it impossible—because of its practices—to achieve them. It is this type of dysfunction that operates against students. And it is this type of dysfunction that the counselor must not only criticize but work to eliminate.

Summary

A number of school systems today are large, impersonal, and bureaucratic in nature. Students in such schools view themselves as victims of the systems with no one to act in their behalf, not even counselors. School counselors, too, have succumbed to the bureaucratic game. Through their unavailability or their strict adherence to a counseling philosophy they have denied students the professional assistance and strategic intervention so desperately needed.

To be relevant today, a counselor must act. It is important that he be a doer. Like the ombudsman, he must harness the power of the bureaucratic school so that it does not overwhelm his main reason for being—the student.

There is much about the ombudsman's role that is relevant to the school. Already colleges and universities are experiencing success with its application there, and there is reason to believe that this same kind of positive response could result from its adaptation to school personnel, especially the school counselor.

The important aspects and functions of the ombudistic counselor's role are *nobility,* or standing between the student and the school; *advocacy,* or advocating needed school changes; *acting,* or doing what is necessary to help the client; and *criticizing,* or "telling it like it is," to improve rather than weaken the system.

References

Abraham, H. J. The Danish ombudsman. *Annals of the American Academy of Political and Social Science,* 1968, **37**, 55–60.

American School Counselor Association. Relevant. *School Counselor,* 1969, **16**, 327.

Bainbridge, J. A civilized thing. *New Yorker,* 1965, **40**, 136–154.

Bexelius, A. The ombudsman for civil affairs. In D. C. Rowat (Ed.), *The Ombudsman: Citizen's Defender.* Toronto: University of Toronto Press, 1968.

Doolittle, L. The concept of ombudsman: its application to teacher education. *Journal of Teacher Education,* 1969, **20**, 160–167.

Jackson, P. W. *Life in Classrooms.* New York: Holt, Rinehart & Winston, 1968.

Noyes, K. J., and McAndrew, G. L. Is this what schools are for? *Saturday Review,* 1968, **69**, 58–59.

Rowland, H. The campus ombudsman. *Today's Education,* 1969, **58**, 37–39.

8.3 A Work Experience Program for the Educable Mentally Retarded at the Elementary School Level

MARY TURNER

Reprinted from *Education and Training of the Mentally Retarded,* 1968, **3,** 199–201, with permission of the author and the Council for Exceptional Children. Mary Turner is a teacher of the educable mentally retarded in the public schools of Santa Cruz, California.

Some students need more attention than others if they are to develop their powers to the fullest capacity. The following article describes one type of program that has been particularly helpful for educable mentally retarded children. It is interesting to note that the program succeeded because it followed sound educational principles: students were involved in an activity that had intrinsic interest for its own sake, and the curriculum was adapted to take advantage of the experiences they were having. Not only did the students become involved in an experience in which they learned valuable work skills, but classroom assignments took on new meaning for them.

In the autumn of 1963 the principal of Westlake School, who is also the supervisor of Special Education in the Santa Cruz City Schools, invited two boys from the intermediate educable mentally retarded (EMR) class to work in the cafeteria at the noon lunch period. From this simple beginning 5 years ago, a work experience program developed which has become the focal point of the intermediate class curriculum. All the academic subjects and other learning activities in this class are related directly to this program.

In the beginning we had no thoughts about vocational training. Although this had been an established part of our high school curriculum for some time, it had not occurred to us that it might be done in the elementary school. It was the enthusiasm of the two boys for their jobs, and the yearning of the rest of the class for similar jobs, that gave the teacher the idea which later developed into this Job Training Program.

Availability of Jobs

During the spring of 1964 the teacher and the principal talked about the possibility of finding jobs for everyone in the class. We agreed that only valid work would do. Unless the job was necessary to the daily functioning of the school plant, it was useless for our purpose. We believed these children could appreciate and benefit by the dignity of honest work. The jobs were there, we were sure. Our problem was to persuade potential employers—the teachers and uncertified personnel—to give our retarded children a chance to prove themselves.

Three classes of younger EMR children are bussed to the school, which is situated in an area of middle to high income homes. Although the EMR children share the same buildings, playgrounds, and recesses with 400 regular class students, few of the teachers or other employees had had close contact with the educable mentally retarded. Understandably they had anxieties about them and misconceptions about their ability.

In spite of this, when classes began in September we had managed to find jobs for 13 children. Some were long term jobs such as handling the dishwasher in the cafeteria, setting tables for lunch, and weeding the newly landscaped grounds. Others were short term jobs lasting only a day or two, such as stacking chairs, assembling and stapling bulletins, painting signs, and hanging out the clothes on the Lost and Found line.

Setting the Goals

Before the jobs were assigned to the children, there was a week of orientation. It was important the children should understand the seriousness of our venture. We discussed what it meant to have a job and be a good worker. At the end of the week we wrote down six ideas, and these became our Rules for Workers, as follows:

1. Be neat and clean
   Have clean skin, hair, teeth, fingernails, clothes
2. Be quiet on the job
   Don't disturb other people by loud talking or banging tools around

3. Be dependable
   Do what the boss tells you even if she's not around
   Don't take anything that belongs to someone else
   Never leave a job unfinished without telling the boss
   Obey all rules even if no one is looking
4. Use good manners
   Don't say or do anything to hurt another's feelings
   Remember Please and Thank you
   Try to be cheerful on the job
5. Be a hard worker
   Don't play around on the job
   Keep your mind on what you're doing
6. Keep good records of your work
   Be accurate

All free lunch tickets were cancelled. From then on it was necessary to work in order to eat. It was also agreed among us that if a worker broke any of the rules, he would be fired and his "pay" stopped. Until he was rehired he must bring his lunch from home. We notified parents of our program and asked their cooperation.

Many children were fired many times that first year. One or two even quit when the glamor of having a job wore off. But the loss of prestige, even more than the loss of the lunch, was immediately felt. There was for the first time a noticeable effort on the part of the unmanageables to improve themselves. The teacher was also learning by trial and error how to help them succeed, so that now it is seldom necessary to fire a child. Even the laziest among us will choose to work rather than suffer the loss of face unemployment causes.

All jobs begin and end between 11:00 and 1:00 each day. It is important to the successful functioning of this program to have everyone at work more or less at the tame time. Each child is expected to keep a daily record of his work in a special book he prepares himself at the beginning of each month when new jobs are assigned. He must record the name of the job and the time he leaves and returns, and add comments about any change in the work routine. In recent years a point system has been established to give credit for accuracy in recording, as well as for any comments made beyond the minimum requirements. These points can be cashed in at the end of the month for desirable goods, chiefly food. To receive points, the child finds he must master such things as reading and writing time, sentence structure, punctuation, and the use of the dictionary in which to locate the words for his sentences.

Very recently we have begun to "pay" the children in play money when their work is finished each day. The money is kept in individual bank boxes. When enough "dollars" have been accumulated, the bank is taken to the office where the secretary requires each child to count out the sum accurately before exchanging the play money for a month's lunch ticket. With this system to help us, the teaching of money concepts has become relatively easy.

Evaluation

Over the years the children's ability to work has been accepted by the staff. The number of available jobs has now increased so that we have begun to

have more jobs than children, and the best workers must sometimes handle two or even three jobs during one day's work time. Among the adults who have given us loud and genuine praise are the school secretary, the manager of the cafeteria, and the Santa Cruz City Schools' librarian whose office has been located in our building. For these women the children have worked hard and earnestly in jobs both simple and complex and in some cases have performed them as well or even better than the children from the regular classes.

When we have visitors, especially those who have worked with mentally retarded children, they are most impressed by our children's social ease. This ease is a result of their contact with many people as they go in and out of classrooms, the office, the cafeteria, and the library on their daily jobs.

Recently we had another kind of proof of our success: some members of our class of 1964 have just finished the beginning phase of the High School Work Experience program. In a written statement the director of this program has commented on the marked differences in readiness for work between our children and those coming from other areas without this background training.

There are so many exciting possibilities yet to be explored, so many techniques to be reshaped and refined, that our Job Training Program will go on growing and expanding and changing almost indefinitely. The enthusiasm and pleasure the children take in the program, and the personal growth we see in them because of the training it provides, bring us immeasureable satisfaction.

8.4 Educational Disability in Emotionally Disturbed Children

F. BETH STONE AND VINTON N. ROWLEY

Reprinted from *Exceptional Children*, 1964, **30**, 423–426, with permission of the authors and the Council for Exceptional Children. F. Beth Stone was a school psychologist in the public schools of Des Moines, Iowa, and is now in Hawaii. Vinton N. Rowley is a clinical psychologist in the Department of Pediatrics of the State University of Iowa.

Emotional disturbances and learning problems go together. In this survey, Stone and Rowley report findings that suggest that the great majority of emotionally disturbed children who are referred for treatment are retarded in reading and arithmetic. Sometimes a child's difficulty in keeping up with his class provides the first clue that he is having emotional problems.

The commonly held assumption that emotionally disturbed children, as a group, have educational disabilities was examined by Tamkin (1960) and shown to be untrue, at least for his sample. He reports that of 34 children admitted for residential treatment at an Ohio psychiatric hospital only 32 percent demonstrated some degree of educational disability, as assessed by the reading and arithmetic parts of the Wide Range Achievement Test, while 41 percent were educationally advanced and 27 percent were at their expected grade level. A possible weakness of his study, which Tamkin himself points out, was "the use of chronological age rather than mental age as the basis from which to measure departures in educational achievement" (p. 314).

The present survey was done to determine whether a large sample of emotionally disturbed children would demonstrate educational disabilities, using both mental age and chronological age as bases for determining departures in educational achievement. A second purpose was to examine the relationship between arithmetic and reading skills in emotionally disturbed children. Tamkin found that his sample had significantly lower scores in arithmetic than in reading.

Method

The subjects consisted of 116 children referred to the Child Psychiatry Service of the State University of Iowa for diagnosis and possible treatment of emotional disorders. There were 82 boys and 34 girls (this proportion of boys to girls is roughly equivalent to the proportion in Tamkin's sample); the mean age was 12.0 years (the mean age in Tamkin's sample was 9.0 years). None of these children had serious medical problems or psychiatric symptoms which required isolation or confinement. The arithmetic and reading parts of the Wide Range Achievement Test (WRAT) were given to each child individually by a psychologist, either on the child's first out-patient visit or during the initial period of his hospitalization. Intelligence quotients were available for all subjects; for 98 of the subjects these were Full Scale IQs on the Wechsler Intelligence Scale for Children (WISC), and for the remaining 18 these were prorated IQs based on five subtests of the WISC which have been found to correlate highly ($R = .96$) with the Full Scale (Enburg, *et al.,* 1961). The intelligence test was administered on the same day as the WRAT in the large majority of cases; in no case did the interval between administration of the WRAT and the WISC exceed three months. The mean IQ of the sample was 96.52 (*sd* 16.39; range 62 to 135).

Results

Educational disability was evaluated on two different bases. To replicate Tamkin's study, achievement grade ratings (i.e., the WRAT grade scores) were subtracted from the chronological ages of all subjects. Remainders lying between 4.83 and 5.83 indicate a level of achievement commensurate with the chronological age since, in Iowa, a child normally enters school (kindergarten) between the ages of 4 years 10 months (4.83 years) and 5 years 10

months (5.83 years). (Tamkin considered remainders between 5.0 and 5.9 as indicative of expected achievement. Different standards are appropriate, since laws governing the age of school eligibility differ from state to state.) Accordingly, remainders above 5.83 would indicate the presence of educational disability, and those below 4.83 would indicate educational advancement.

The mean difference between chronological age and grade rating in arithmetic was found to be 7.21 (*sd* 2.10) and in reading, 6.47 (*sd* 2.49), suggesting that the achievement of this sample of emotionally disturbed children is not commensurate with their chronological age. The distribution of cases falling in the three levels of achievement—disabled, expected, and advanced in terms of chronological age—on both arithmetic and reading is presented in Table 1. In both reading and arithmetic, the majority of children in this sample fell below the level of achievement expected on the basis of their chronological ages.

A similar analysis was done using mental age as the basis for determining achievement level. That is, a subject's grade ratings on the arithmetic and reading tests were subtracted from his mental age. A remainder between 4.83 and 5.83 indicates a level of achievement commensurate with mental age, a remainder above 5.83 indicates achievement below that expected on the basis of mental age, and a remainder below 4.83 indicates achievement above that expected for mental age. Since the WISC is not constructed to yield a mental age, the subjects' mental ages were obtained by the conventional formula. No child whose mental age was less than 5.83 was included in this sample since the remainder, MA minus grade rating, would never be greater than 5.83 and hence would lead to the child's being classified as "expected" or "advanced." Only three children out of an initial sample of 119 had mental ages below 5.83.

The mean difference between mental age and grade rating in arithmetic was found to be 6.69 (*sd* 1.51) and in reading to be 5.94 (*sd* 2.08). Table 2 shows the distribution of cases in the three levels of achievement determined on the basis of mental age. Although the proportions shifted upward, the majority of children still fell below the ex-

TABLE 1. Distribution of 116 Emotionally Disturbed Children by Achievement Levels Based on Chronological Age

	Reading		*Arithmetic*	
Level	Frequency	Percent	Frequency	Percent
Disabled (mean remainder $\geq$ 5.83)	69	59.48	86	74.14
Expected (mean remainder $<$ 5.83, $\geq$ 4.83)	24	20.69	20	17.24
Advanced (mean remainder $<$ 4.83)	23	19.83	10	8.62

TABLE 2. Distribution of 116 Emotionally Disturbed Children by Achievement Levels Based on Mental Age

Level	Reading		Arithmetic	
	Frequency	Percent	Frequency	Percent
Disabled (mean remainder $\geqq$ 5.83)	60	51.72	78	67.24
Expected (mean remainder $<$ 5.83, $\geqq$ 4.83)	22	18.97	27	23.28
Advanced (mean remainder $<$ 4.83)	34	29.31	11	9.48

pected level in both arithmetic and reading, even when achievement levels were based on mental age rather than chronological age.

Since performance level on an achievement test must be partly a function of actual grade placement, the actual grade placements of the children in this sample were examined. It was found that 47.83 percent of the children were in grades that would be expected on the basis of chronological age, 43.48 percent were in grades below that expected (including those children, six out of 116, who were placed in special education classes), and only 8.70 percent were in grades above that expected.

As in Tamkin's study, the emotionally disturbed children in this sample had lower arithmetic scores than reading scores. The mean grade rating in reading was 5.49 (*sd* 2.64), and the mean grade rating in arithmetic was 4.74 (*sd* 2.21), this difference being significant at the .001 level ($t = 4.746$).

Discussion

Although in Tamkin's sample, educational disabilities did not predominate, the present results indicate that in a larger, more heterogeneous (in terms of age) sample of emotionally disturbed children, educational disabilities do predominate, even when mental age is the basis from which departures in achievement are figured. Tamkin concluded that his findings "lent support to the theory that educational disability may be a symptom of the same condition producing the emotional problems rather than being caused by the emotional problems" (p. 314). While the present results are not in agreement with Tamkin's results, they do not necessarily refute his theoretical position. The kind of evidence needed to test such a hypothesis is not provided by a survey, in the present authors' opinion. Neither a survey of the educational disabilities of a sample of emotionally disturbed children nor a survey of the emotional problems of a sample of educationally disabled children provides sufficient evidence to test a hypothesis about the causal relationships between emotional problems and learning disability.

Jastak (1946) opined that arithmetic disability is often indicative of severe personality difficulties and is more fre-

quently associated with maladjustments of a personal nature than reading disability. Tamkin views his finding that emotionally disturbed children score lower on arithmetic than on reading as supporting Jastak's opinion. Although the present results are in agreement with Tamkin's results on this point, they do not necessarily support his theoretical position. There may be other factors, unrelated to personality factors, which contribute to a discrepancey between arithmetic and reading scores. A child's actual grade placement—a factor not taken into account by Tamkin—has a differential effect on reading and arithmetic achievement, the child's achievement in arithmetic being more limited by the classroom instruction he has (or has not) received than his reading achievement (Stroud, 1956, p. 377). Hence, it is possible that a group of emotionally disturbed children is more disabled in arithmetic than in reading simply because many of them are behind in actual grade placement. The fact that 43 percent in the present sample were behind suggests that this is one factor that must be taken into account before the discrepancy between arithmetic and reading can be viewed as a function of personality factors. Again, in this regard, a survey does not provide sufficient evidence to draw conclusions about the causal relationships between arithmetic disability and personality disorganization.

A question that might be raised in regard to the results of both this study and Tamkin's study is whether the "expected" achievement level is not too narrow. Can one reasonably expect the majority of any group of children to fall within a one year range of achievement? A more basic question, however, is, "What is the expected range of variability of WRAT scores of children in the general population?" Data on this question do not appear in the test manual (Jastak, 1946), nor have they been published elsewhere, apparently. These surveys point out the need for more normative data on the WRAT.

One might expect that the children in the present sample, since they are of a wider age range, would have more variable achievement scores than the children in Tamkin's sample. It seems quite probable that WRAT scores become more variable as age- and grade-level increase. Unfortunately, Tamkin did not indicate the variability of WRAT scores for his sample, nor have data been published on the variability of WRAT scores for unselected samples of school children in different age- and grade-levels. Nonetheless, even greater variability in the present sample would hardly account for the greater proportion of educationally disabled children.

Summary

A survey of 116 emotionally disturbed children revealed that educational disabilities in both reading and arithmetic predominated when mental age, as well as chronological age, was used as the basis from which departures in achievement were measured. It also revealed that in actual grade placement a much larger proportion were in grades below that expected on the basis of chronological age than above. Arithmetic scores were found to be significantly lower than reading scores. The limita-

tions of survey data in drawing conclusions about the causal relationships between emotional problems and educational disability, as well as the need for additional normative data on the Wide Range Achievement Test, are discussed.

References

Enburg, R., Rowley, V. N., and Stone, Beth. Short forms of the WISC for use with emotionally disturbed children. *Journal of Clinical Psychology,* 1961, **17**, 280–284.

Jastak, J. *Wide Range Achievement Test, manual.* Wilmington, Delaware: Charles L. Story, 1946.

Stroud, J. B. *Psychology in Education.* New York: Longman's, Green, 1956.

Tamkin, A. S. A survey of educational disability in emotionally disturbed children. *Journal of Education Research,* 1960, **53**, 313–315.

8.5 *School Grades and Group Therapy*

SAMUEL TENENBAUM

Reprinted from *Mental Hygiene*, 1970, **54**, 525–529, with permission of the author and the National Association for Mental Health, Inc. Samuel Tenenbaum is professor of counseling and guidance at Long Island University, and has written extensively in the associated fields of mental health and education.

In the preceding article, Stone and Rowley found evidence pointing to a relationship between emotional problems and school difficulties. There are probably elements of mutual causality involved in this relationship, and improvement in one ought to be accompanied by improvement in the other. In the present article, Tenenbaum reports what happened when students' problems were relieved through group therapy. Although the setting of his small experiment is the university, his method and its results are equally applicable to secondary schools.

There are no lazy people. There are frightened people, anxious people, bored people who have found no meaning in life and hence no meaning in the activities necessary for life. And there are people who have unrealistic notions of who they are and what they are. They, too, are crippled for realistic living. As for students, what keeps them from their work for the most part is fear of failure, terrible self-demands, low self esteem, inability—intellectually or emotionally—to cope with the tasks set before them. In our culture, students find it more acceptable and more respectable to plead disinterest and laziness, rather than lack of capacity and intelligence, and this is the way most students and parents prefer to explain failure.

In the main, young people who have school problems are immobilized by their fears. When a child is doing poorly at school, teachers are wont to attribute it to indolence and laziness; rarely to his fears, his hang ups or his great desire to please his parents, all of which may be debilitating and self- defeating.

There were a number of education major students at Long Island University who had been on probation semester after semester. Before expelling them, the then Dean of the Division of Education, and the adviser of the Educational Counseling Service thought that as a last chance, they ought to be given the opportunity of having a group therapy experience. I was asked if I would like to take over the group and I accepted, having advocated such group experiences, not only for students in academic difficulties, but for all our students for many years.

About 15 students were originally assigned to the group. Six came once or twice and never came back. I was left with a group of nine, six of whom came regularly and three irregularly.

In my private practice I seldom experience a cancellation. The thought that students who did not pay a fee would reject the chance for a therapeutic experience was something I could not contemplate without of severe blow to my self-esteem. But this was my own problem and I never communicated this to the group. I held firmly to the belief (and I still do) that it was for each student to decide whether he wanted to attend or not; and if a member stayed away, I interpreted it to mean that he wasn't profiting and it was, then, a reflection on me and not on him. Every absence made me feel very uneasy and guilty.

From the outset I never asked students to work or study harder, for I assumed that the thought had occurred to them; or if not, someone else had told them of the need. Never once did I bring up the matter of grades unless they did. I did try to listen, to be sensitive to them as they expressed their grief, worry and often despair. As I listened I often wished that faculty members could be present and hear these anguished communications, so that they could understand what a grade means.

Although students could attend at most ten sessions of one hour and a quarter, not one of the nine failed a single course, not only during the se-

mester in which they participated in the group experience but also in the semester that followed. Where hitherto their records had been filled with Fs, Ds and Cs, after their group experience there were several As, B+s, Bs and C+s.

Prior to their group experience, grades C+ and higher for these students came to 24.7%; after the group experience it rose sharply to 65.0%. Before the group experience, grade C and below came to 76.3%; after, it was 34.5%. In all their previous school experience, these students had failed 10.4% of their courses. After their group experience, they did not experience a single failure or F grade for the entire academic year. As for the D grade, there was a sharp drop, from 31.2% to 9.6%. Even for the gentlemanly pass, the C grade, there was also a drop from 28.9% before to 22.5% after. These percentage deficites in C and lower grades were made up by the sharply upward movement to grades above C after the students became involved with the group.

In their prior years at Long Island, (two had attended the university six semesters and seven, four semesters), these students made altogether one A, which was 0.5% of all their grades. In the semester concurrent with and the semester following their group experience, they made eight As, representing 8.6% of all their grades. Before their group experience, their B+ grades came to 2.3%, after, 10.6%; their B grades before, 9.8%, after, 25.8%; C+ 10.4% before, after 17.1%. After the group experience, their grades rose markedly from the lower end to the higher end of the grading scale.

During the semester of the group meetings, I asked the students to write brief reactions to each session and several of those are quoted here.

At the start, they found talking difficult; and their communication was random and desultory. They could not understand how plain, ordinary talk could help them with their grades. All were reluctant to reveal their probationary status and their silences were long and severe.

These are two typical early reactions:

> I noticed today that we, as a group, have a lot in common as to why we are on probation, but that our fears of being with strangers prevent our really coming out with the true facts of our problem. . . . I am sure if someone starts, others will follow.
> I don't see how anyone can really be helped. . . . and there are many things I won't say because there are other people in the room.

One member took to the process immediately:

> After leaving the first session I felt as if a great burden was lifted from my shoulders because for the first time in my life I could speak freely of the personal problems which I have been faced with in the last few years. . . . For the first time in a long while I didn't feel any anxiety about being in a classroom.

Mostly, however, they did not know how to articulate feelings; and they resented having to make the effort. This

was a hard stage for them and for me. Wrote one member:

If it seems that I don't participate as much as the others, it's just that I get very upset telling my problems. . . . I hope at future meetings it won't bother me as much to tell my problems.

Gradually, they talked less and less about school and grades and sought each other out for support and comfort. Bound as they were by a common grief—the pain and the hurt and the shame of their probationary status—the group slowly began to jell and assume a close, in-group quality.

I was happy during this session because I felt that we are learning different facets of each other's personalities. I felt in the past our lives started and ended with classes and that we had no other lives outside of school. During the hour we spoke of ourselves socially.

It is amazing to what extent parents figure as a disturbing influence to the members of the group; how, instead of helping them, they fill them with disturbing emotions, so that the tasks in connection with school do not become a personal student involvement but a way to win parental approval and an honorable and respected position in the family. For these students, school represents nothing but a vehicle to get grades sufficiently high so that parents will think well of them. For the openly rebellious, a struggle may develop between student and parents. The parents will badger their offspring: "Have you studied enough?" "What grades did you get on your examination?" "Have you done your homework?" The educational experience becomes subsidiary to parental needs and demands. In such instances, there may be acrimony and bitterness and sometimes despair, not because of what goes on at school and what school represents, but at what goes on at home. Here are typical student reactions:

I was very glad to see that I'm not the only one with mother problems.

One thing we have in common is meddling parents. I am not saying this is the cause of our probation status, but it may very well be one of the causes.

The parents took the responsibility of the school task away from their children. Further, no matter how they rebelled and how bitter the acrimony, these students took it for granted that their parents were right and that they were worthless. Although this did not make them better able to cope with the school situation, it did make them feel guilty and inadequate. These feelings, in my judgment, only accentuated students' anxieties and made them less able to meet school demands. In the group sessions, students began to perceive how parents figured in the school situation, and to understand better the nature of their parental relationships. One student writes:

I realize that my mother will never be pleased. That is something I never realized. I'm sure this . . . will help me . . . now and in the future. For if I keep listening to her, I may find I'm living for her rather than for me.

Although the matter of grades and probation was always present, the discussions soon went far afield of grades. They included their whole persons; how they struggled, how they failed, also, how they triumphed. In my opinion, this proved to be the most valuable part of the experience since it helped them gain realistic insights into their own difficulties and their own persons.

I feel that much was accomplished (in our last session) in that the meeting really brought me to thinking why I got marks as I did. . . . The family situations described (by others) were somewhat like my own so I was helped in that area as well. . . . Though I don't fully understand why I've been so erratic in my work, I know that it has been at least 90% of my own doing and I aim to better myself.

In the final stage, they were a united group, working in unison, understanding one another, better aware of their own feelings and their own problems, more realistic about what needed to be done. Below is one reaction articulating this new awareness of self, these new insights and new resolutions.

I truly believe that our open-mindedness about our individual problems has helped us to achieve our success. If we can continue to verbalize about ourselves, then surely we can allay our fears about school. . . . Of course, we have to work and study, but I feel that by talking, half the battle to get off and stay off probation is won.

I asked for a final reaction as to how they viewed their experience and requested that they send me their final grades. These were mailed to me after their final examinations and long after the group had broken up. I wish there was sufficient space to quote these communications verbatim.

One in particular sums up the many reactions:

. . . it was a very good feeling to know that there were others in the same category as myself; and that there was a person, such as yourself, who seemed to take an honest interest in each individual. I feel that now that I am off probation, I will stay off.

Discussion

At first the group members did not know how to talk and were reluctant to talk. Each member of the group thought that "being on probation" was unique to him and each carried this burden for the most part secretly and in shame. When they discovered that all in the group were on probation, it had an exciting, liberating effect. Eventually, their common problem served as a bond and a tie to unite them.

I want to emphasize that in none of the sessions did I bring up the matter of their grades; in no instance did I make them feel that I wanted them to get higher grades, or that I was in any way involved with their grades.

Although the group members' probationary status was always present, once they revealed themselves as persons, they discussed grades from a deeper and more significant point of view. Grades became linked with their anxieties, their hopes, their weaknesses, their failures, their parents, etc., etc.

In the group, the members received regard, encouragement, support, affection. Their ego strength increased, their self-respect rose, and as one of the outcomes, I believe, they were able to confront their school situation in a healthier and more intelligent way. They were able to view themselves and their problems more insightfully, more realistically. One spoke of anxieties so great that she could scarcely live through an examination. Others spoke of frightening instructors who marked you down for anything and everything so one couldn't think. My heart went out to these unfortunates, as they tried so hard to make themselves small and inconspicuous in class, fearful of being called on. Others spoke of personal situations which they faced that kept them so distraught that they were immobilized from doing anything. "I got started studying and then my mother got after me and we had the worst fight in a month and I couldn't study, and she said I was no good and I would never be good." The tears rolled down her eyes. "I couldn't study . . . and I couldn't sleep that night."

The group experience made the members more articulate. Before they were silent in class—outsiders. As they became more aware of their feelings and better able to articulate them, several managed to develop more personal relationships with some of their instructors and, best of all, a number became more active as class participants. These began to feel like persons and act as persons, not nonentities.

When I discovered that the group members made even higher grades in the following semester than they did while they were undergoing the group experience, I was puzzled until the thought occurred to me that they did not have the full force of whatever is therapeutic in a group until the following semester; and that this could account for the difference.

Although they were doing unsatisfactory school work, it appears that every member of the group had the requisite intelligence to perform the academic tasks set before them. In fact, several became, after the group experience, not only successfully functioning students but superior students. The question arises: On what exactly does academic success and failure hinge? Are grades a symptom of the whole functioning person, who and what he is, and not primarily a matter of intelligence which hitherto was regarded as the main and only component. If yes, what are these other factors? Further, if these students had not undergone the group experience, what would have happened to them? How would their lives have been changed? This is a matter not to be taken lightly. In our society, college graduation is the union card for valuable rights and prerequisites, vocational, social and even marital. Would they have righted themselves on their own or would they have been expelled as academic failures with all its inherent emotional and psychological trauma, very often lifelong.

This is not a statistical study. It is rather an account of nine persons, each one apart and separate, who participated in a group experience. Although we can conjecture, I do not believe that at this stage we know exactly

what happens to the members in a group of this kind; how a group affects each in it; and how each in it affects the group.

The hope is that with time, with further study and inquiry, the process will be better understood and therefore better controlled, and hence, we will be more able to replicate outcomes. Even with our present knowledge, we have every reason to believe that when a group works, it can be highly salubrious and therapeutic, in ways which we have not nearly plumbed.

8.6 *New Programs for Dropouts*

HERMAN SLOTKIN

Reprinted from *Vocational Guidance Quarterly,* 1963, **12,** 127–132, with permission of the author and the National Vocational Guidance Association, Inc.
Herman Slotkin is Coordinator of Job Education Projects for the New York City Board of Education. He also teaches guidance at New York University and has a special interest in the problem of dropouts and underachievers.

Both employment and schooling possess the advantage of providing opportunities for learning and for potentially ego-building, therapeutic experiences. Most young people of school age find these rewards in schooling, at least to some degree, but some become discouraged with continued failure at school and seek success in employment. Many of those who drop out, however, encounter the same problems at work as they did at school; others are somewhat more successful, but still are not fulfilled by their work. The solution to the problem of the potential dropout seems to be that of working out a combination of employment and school experience that is complementary and mutually rewarding. We noted how this can be done with educable mentally retarded children in the article by Turner; this report by Slotkin tells of successes attained by work-study programs with socially disadvantaged students in New York City.

In the Summer, 1962 issue of *The Vocational Guidance Quarterly,* an article described three New York City experimental projects designed to prevent early school leaving and to rehabilitate those who left school before graduation. The evaluation of these experiments was completed in December, 1962. The results were so encouraging that the Board of Education has incorporated all three projects in the educational offerings of the City and has made budgetary provision for their expansion throughout the system. This article presents the findings of the evaluation.

The three experiments were conceived as a coordinated attack on major aspects of the dropout problem. Project I provided a study, work, and guidance program designed to retain potential dropouts until graduation; Project II was an evening school program for school-oriented dropouts; Project III, a program of pre-employment preparation, job placement, and followup, was intended for school-alienated work-oriented boys and girls.

In all of these programs the growth of actual or potential dropouts toward clearer, more realistic vocational goals was regarded as central to motivating achievement at school and at work.

Evaluation Plan

The projects were started in February, 1961. The target date for the conclusion of the experimental period was September, 1962, and the projected date for the conclusion of the evaluation was June, 1963. The experimental designs involving matched experimental and control groups were developed with these termination dates in mind. Followup of students beyond a period of about 19 months was therefore not feasible for this report.

The matching criteria were designed to equate experimental and control groups with regard to a number of significant factors:

1. To equalize school, neighborhood, and other environmental conditions, experimentals were matched with controls who left (or planned to leave) the *same school* at the *same time.*
2. To eliminate differences that might accrue from differences in ethnic origin, experimentals and controls were matched with regard to this factor.
3. To minimize the differences in maturation and eligibility for various kinds of work, both groups were equated with regard to chronological age.
4. To eliminate differences in scholastic aptitude and reading ability, experimentals and controls were matched with regard to IQ and reading grade. The intelligence test used in all but a few cases was the Pintner General Ability Test, Intermediate, and the reading test was the Stanford Achievement Test, Advanced Reading, both given in the 8th grade.
5. To equalize the experimentals and controls with regard to actual school accomplishment, they were matched with regard to school achievement (majors passed divided by majors taken).

6. To control for sex, pairs were matched on this factor.

Each experimental student was matched with a control student within specific limits on each criterion. The groups were then checked with regard to central tendency and dispersion to verify the match. Depending upon the availability of data, the number of matched pairs for each criterion comparison differed. For example, in Project III, in a comparison with regard to the job-performance criterion, there were 67 pairs; in the comparison of the total hours of school instruction criterion, there were 103.

Wherever it was appropriate and feasible, other matching criteria were introduced and utilized so that experimentals and controls have, in some cases, also been matched as to rate of school attendance and ratings of personal-social traits, as well as other characteristics.

It was recognized that in a complex school situation, no matching could be perfect, but, insofar as was possible, major relevant factors were accounted for.

In comparing experimental and control groups, the criteria for evaluating the success or failure of the school programs fell into four categories:

1. A direct measure of the attainment of one of the prime objectives of the programs, "the attainment of a clear, realistic, vocational goal" by the pupils in the experiment.
2. Measures of the results of the new programs in terms of school achievement and adjustment.
3. Measures of the results of the new programs in terms of job achievement and adjustment.
4. Measures of attitude toward the various school programs on the part of students, parents, employers.

Information regarding school achievement and adjustment was gathered from school records. Information regarding the other three criteria for evaluation was gathered by means of separate questionnaires sent to student, parent, and employer of experimentals and controls at the same time, after the start of the experiment. The questionnaire information was supplemented by interview data secured by counselors from students, parents, and employers.

The results of the evaluation are summarized below.

Findings for Project I

In Project I, a study, work, and guidance program of dropout prevention, the primary goal was the retention of the experimental boys and girls in school until graduation. Failing this, it was a purpose to prepare for improved achievement and adjustment at work.

The 53 experimentals and 53 controls were drawn from a group of 180 youngsters in one high school who exhibited the prime signs of early-school-leaving: they were academic failures; they were truant; they were discipline problems. Each experimental boy and girl was matched with a control as to age, IQ, reading, school achievement, sex, ethnic origin, progress grade, attendance, and teachers ratings of personal-social characteristics.

The control group of potential dropouts received the normal education and services available in the school. Each experimental student was assigned a new major subject, Pre-employment Guidance, in which he assessed his interests, abilities, and aptitudes in terms of job demands in the current labor market. It was hypothesized that group guidance coordinated with individual guidance and testing, and focusing on vocational development, would lead each student to a clearer more realistic vocational goal. This, in turn, would lead to improved achievement at school and at work.

In addition, experimental students were provided opportunities to participate in remedial instruction, skill subjects, and part-time work experience. Successful and unsuccessful experiences, both in school and out, were evaluated in group and individual guidance to enhance each pupil's understanding of himself.

At the end of a 19 month experimental period, the control boys and girls were compared with the experimentals.

Table 1 presents the major findings of the evaluation of Project I.

In addition, a comparison of the experimental and control groups with regard to the students' attainment of clear, realistic, vocational goals at the end of the experimental period showed that the experimentals were superior in this respect ($P < 0.10$). However, this result failed to meet our test of significance ($P < 0.05$). Counselors rated pre- and post-plans of experimentals and found that two thirds of them had developed clearer, more realistic vocational goals.

Since very few members of the experimental group had left school by the cut-off date for the collection of data, it was not possible to compare experimental and control groups as to job achievement and adjustment. A second followup is planned for April and May of 1964 when all subjects will have had some work experience.

Pupils, parents, and employers of experimentals and controls in Project I rated the school programs. A comparison of the ratings showed that more experimentals than controls in each

TABLE 1. Differences between Project I Experimental and Control Groups

Evaluation Criterion	N (in Pairs)	Experimental Group	Control Group	
Retention–graduation rate	53	0.64	0.36	< 0.02
School achievement ratio	53	0.72	0.50	< 0.01
Teachers' rating of personal-social characteristics [a]	31	2.10	2.31	< 0.05
Months in school	53	13.26	9.28	< 0.01
Rate of attendance	53	0.87	0.89	Not sig.

[a] The higher the score the worse the rating.

case rated the school program as "helpful," but only in the case of the employers' ratings did the differences reach the required statistical significance.

In the attainment of its primary goals, high school graduation and improved achievement, the project was successful. Over the experimental period of more than a year and a half, 64 per cent of the potential dropouts in the experimental group either graduated or were still in school. Only 36 per cent of the matched controls were in the same category. In addition, the students in the experimental group were superior in scholarship and in teachers' ratings of their dependability, cooperation, courtesy, and appearance. In effect, the boys and girls in the experimental group graduated in greater numbers, stayed in school longer, and did better school work than they were expected to do, and better than the boys and girls with whom they were matched.

Findings for Project II

The evening schools program, Project II, was designed for 16-year-old boys and girls who left school but were sufficiently school oriented to be induced to go to evening school for a diploma or for the improvement of work skills. A counselor was provided in each of the experimental evening schools to ease the student-worker over the inevitable school and work problems that arise.

The 103 students in this experimental group were paired with 103 boys and girls who dropped out of the same schools at the same time and were matched as to age, sex, ethnic origin, IQ, reading level, school achievement, and attendance. The controls attended the continuation school for one-half day a week. The experimentals took a minimum of two major subjects in evening school.

Table 2 presents the principal findings for Project II.

With regard to the attainment of

TABLE 2. Differences between Project II Experimental and Control Groups

Evaluation Criterion	N (in Pairs)	Experimental Group	Control Group	
Rate of attendance	103	0.70	0.39	< 0.01
Hours per week of instruction	103	5.01	1.52	< 0.01
Teachers' ratings of personal-social characteristics [a]	40	1.89	2.36	< 0.01
Reenrollment rate—full time school	103	0.04	0.01	
School achievement ratio [b]	48	0.90	0.59	< 0.01
Unemployment rate	103	0.10	0.24	< 0.01
Employer ratings: job performance [a]	67	2.82	3.63	< 0.01
Employer ratings: job attitude [a]	67	2.76	3.82	< 0.01

[a] The higher the score the worse the rating.

[b] With regard to this criterion only, comparison was made between pre- and post-performance of experimental cases. No school achievement data were available for the control group.

clear, realistic vocational goals, the experimental group was significantly superior to the control group ($P < 0.01$). Students in the experimental group, their parents, and their employers, rated the school program "helpful" in significantly greater numbers than their control counterparts ($P < 0.01$).

Evaluation showed that in school adjustment and achievement, in the attainment of clear, realistic vocational goals, and in the attitudes of students, parents, and employers toward the school programs, the experimental students proved to be significantly better than the controls. Of particular meaning was the success of the experimental boys and girls at work. Their unemployment rate was less than half that of the controls, and they were rated superior in work performance and attitude by their employers.

Findings for Project III

Project III was intended for the most difficult teenagers—those who are alienated from school and are lacking in employability. Specifically it was the purpose of this project to prepare students for employment in one month, help students to begin the process of attaining clear, realistic vocational goals, assist in the students' job adjustment, and return as many as possible to the regular stream of education. Again, the central concept was that the attainment of a clear, realistic vocational goal is dependent on assessing oneself in terms of the real demands of work; the initial hypothesis was that attainment of a clear, realistic vocational goal would lead to superior achievement at school and at work.

To carry out these aims, experimental students were given a preemployment course lasting twenty school days. The first days of the course were devoted to finding the interest, aptitudes, and skills of each student. Group and individual guidance supplemented by standardized tests were used in the process. The last part of the course was used to review employment opportunities, to sharpen job getting skills, and to develop one or more of these minimal skills for an entrance job: filing, stock keeping and inventory, operating an adding machine or cash register, use of a postal scale, use of the telephone, merchandise marking and display, packaging and bagging, and, for some, typing.

At the end of the course each experimental student was placed in a job by a New York State Employment Service counselor. A school counselor visited or contacted each place of employment at least once a month to check on the work adjustment of the new workers. Those who were found to be unemployed were returned to school for additional instruction or another job referral. It was expected that patient and persistent evaluation by student and counselor of each unsuccessful experience augmented by appropriate instruction would lead to successful job experience.

In the evaluation of the effectiveness of this project, each of the 151 experimental youngsters was paired with a child who dropped out of the same school at the same time. Each pair was

matched as to IQ, reading, grade, age, sex, ethnic origin, school achievement, and teachers' ratings of personal-social characteristics. The experimental group participated in the program we have described; the controls received the conventional school program for 16-year-old dropouts, one half day per week of instruction and service in a continuation school.

Table 3 presents the major results of the comparison between experimental and control groups.

In addition to the findings presented in Table 3, the experimental group was significantly superior to the control group in the achievement of clear, realistic vocational goals ($P < 0.01$). Also, students in the experimental group, their parents, and their employers rated the school program "helpful" in significantly greater numbers than their control counterparts ($P < 0.01$).

The findings of the evaluation showed that in school adjustment, in the attainment of clear, realistic vocational goals, and in the attitudes of students, parents and employers to the school programs, the experimental boys and girls were significantly better. Moreover, these youngsters, who were deemed alienated, returned to full-time school in greater numbers than did their matched control counterparts. In addition, despite the fact that they came to the program as unemployable, they were more successful at work. Their unemployment rate was half that of the controls, and they were rated superior in work performance and attitude by their employers.

Some Implications

(1) In dealing with the dropout problem and manpower retraining there has been a tendency to concentrate on the expansion of educational and vocational opportunities with the expectation that these opportunities would be self motivating. The results obtained in Projects I, II, and III suggest that for

TABLE 3. Differences between Project III Experimental and Control Groups

Evaluation Criterion	N (in Pairs)	Experimental Group	Control Group	
Rate of attendance	151	0.88	0.38	< 0.01
Hours per week of instruction	151	5.11	1.54	< 0.01
Teachers' ratings of personal-social characteristics [a]	42	1.94	2.57	< 0.01
Reenrollment rate—full time school	151	0.13	0.01	< 0.01
Unemployment rate	151	0.17	0.32	< 0.01
Employers' ratings of job performance [a]	72	3.01	3.73	< 0.01
Employers' ratings of job attitude [a]	72	3.10	3.86	< 0.01

[a] The higher the score, the worse the rating.

self motivation to occur, it is equally important to focus on the vocational development of each individual. The development of new educational and vocational opportunities must be accompanied by attention to the vocational development of those who enter these opportunities to produce the best results.

(2) If careful attention to the vocational development of dropouts and potential dropouts can improve their achievement at school and at work, it is suggested that similar treatment for all underachieving students may effect some improvement in their educational attainment.

Shelly Rusten

PART 9

The Socially Disadvantaged Learner

9.1 *Malnutrition and Learning*

RITA BAKAN

Reprinted from *Phi Delta Kappan,* 1970, **51,** 527–530, with permission of the author and Phi Delta Kappa, Inc. Rita Bakan is on the faculty of the Center for Urban Affairs of Michigan State University.

The first thing that comes to mind when one thinks of poverty is hunger. It would seem obvious that the children of the very poor would be less well nourished than other children and, as a consequence, would have more problems with school. That poverty children have more school problems has been well established, but it has been easier to relate them to psychological and social factors than to malnourishment. No one wants to put children deliberately on a reduced diet to conduct a controlled experiment in order to determine the relative effect of psycho-social variables and nourishment on school progress. Common sense has therefore prevailed, and subsidized lunch programs were introduced in schools many years ago as a partial solution to an unresearched problem. In the meantime, research data has accumulated from a variety of sources, showing that nutrition level does have a significant effect on cognitive functioning. They are summarized in this article by Bakan.

To paraphrase Hemingway's dialogue with Fitzgerald, the poor are different from us—they have less money.[1] They also have children who in general do poorly in school and on tests of intelligence. In part this is due to social and psychological factors, and such children have been called "culturally disadvantaged" or "socially deprived." But the poor also have a higher infant mortality rate, a higher incidence of infectious and chronic diseases, and a greater number of premature and low birth-weight infants.[2] All of these factors are in turn related to the disproportionate number of poor children who are mentally retarded.[3]

Data currently being developed indicate that malnutrition is the common denominator of all these ills. The characteristics of malnourished children bear a striking resemblance to a number of the known characteristics of "disadvantaged" children, e.g., apathy, irritability, sickliness, and a reduced attention span.

The combination of malnutrition and the other negative effects of poverty perpetuates a cycle of illness, educational failure, and more poverty. If the aim of compensatory education is to break this cycle, then it is imperative that social scientists and educators become aware of the negative impact of malnutrition on the development of the child. The purpose of this paper is to present some of the evidence, based on animal and human studies, linking malnutrition to the growth of the brain, the performance of various intellectual functions, and other developmental variables related to learning.

Brain Development and Function

Numerous investigations[4] of the effect of malnutrition on the weight of the brain of the rat and of the pig have reported similar results: Undernutrition from birth to 21 days produces a persistent and permanent reduction in brain weight. The earlier the malnutrition the more severe is the effect and the less likely is recovery. Undernutrition also results in specific degeneration within brain cells; again, the earlier the restriction, the more severe the damage.

Results of studies of malnutrition on cellular growth patterns indicate that early restriction of food interferes with cell division and that the animal is left with a deficit in the *number* of cells in all organs, even after adequate refeeding. Late malnutrition results in a decrease in cell size, with recovery on refeeding. While the extent and duration of undernutrition and rehabilitation are important, the age at which the undernutrition occurs appears to be the crucial variable.

There is also evidence of what may be termed "double deprivation." The offspring of rats who are malnourished during pregnancy show a reduced number of brain cells. Even if they are fed normally after birth they do not recover and are left with a permanent deficit in brain cell number. The effects of postnatal malnutrition on animals who have already suffered prenatal malnutrition are more marked than effects of either prenatal or postnatal deprivation separately. It would seem that prenatal malnutrition made these ani-

mals more susceptible to postnatal undernutrition.

In general, the animal studies have shown that during the preweaning period in the rat an enormous amount of chemical change is taking place within the brain and that this is the time when the brain is most sensitive to the detrimental effects of undernutrition. The brain of the mature rat who was malnourished during this period is not only physically smaller but shows degenerative cell changes. If the deprivation occurs early in infancy, these changes are irreversible, while the effects of later deprivation may be reversed through proper feeding.

Behavioral changes were also evident as the result of malnutrition. After only four days on a restricted diet, rats began to develop spasmodic trembling of the head and forepaws, pigs exhibited various motor abnormalities, and puppies became hyperirritable.[5] A decrease in exploratory behavior was suffered by rats who were malnourished during the preweaning period.[6]

One of the most intriguing findings in this area is that poor nutrition of the infant female may affect the development of her offspring many years later.[5] On the basis of a number of studies, Cowley[7] suggests that the nutritional history of previous generations must be taken into consideration when studying the present generation. Whether the poorer test behavior of successive generations of animals fed deficient diets was due to enhanced nutritional deficiency in the young or to differences in the behavior of malnourished mothers toward their offspring is yet to be determined.

Available evidence from human studies reinforces the findings of experiments with animals and suggests that early infancy is a critical period for the development of the brain. This is also the time when the brain is extremely vulnerable to the effects of malnutrition.[8] Since direct measurement of brain growth in humans is not possible with living children, an indirect measure which has been widely employed is the rate of increase in head circumference. Malnutrition, especially during the first year of life, will curtail the normal rate of increase in head circumference.[9] This reduced head circumference of malnourished children, during the first six months of life, according to Winick,[10] accurately reflects the reduced number of cells present in their brains. Monckeberg[11] has demonstrated that the brain of a severely malnourished child may be even smaller than the head circumference would indicate. He has developed a "transillumination" test which reveals the presence of spinal fluid in the skull. The test is conducted by focusing a thousand-watt light on the top of the skull. As the light is diffused, the surrounding area glows. In normal children, this area is very small, but in the affected children the entire brain case glows, from the forehead to the back of the head. A fluid, similar to spinal fluid, fills the cavity between the brain and the skull.

What is the relationship of these changes in brain development to behavior and to intelligence?

The children in the Monckeberg study were restored to physical health,

but follow-up intelligence tests showed that they achieved lower scores than children who had not suffered from malnutrition. Their greatest deficit was in the area of language development. Similarly, the malnourished children studied by Stoch and Smythe[12] who exhibited reduced head circumference had lower I.Q.'s even after long-term follow-up.

Using a wide battery of tests, Klein and Gilbert[13] studied the performance of malnourished and normally nourished children from the same social class. On tests where the test stimulus was only available for a short period of time, the malnourished children performed less well than the normally nourished. There was no difference in performance when the duration of the test stimulus was not controlled. The deficit which the malnourished children exhibited seems to involve speed of perception or of information processing and impairment of short-term memory.

A series of studies by Cravioto and his associates[14] in Mexico and Guatemala has shown that performance of children on psychological tests was related to nutritional factors, not to differences in personal hygiene, housing, cash income, or other social and economic variables. The performance of both pre-school and school children in Mexico on the Terman-Merrill, Gesell, and Goodenough draw-a-man tests was positively correlated with body weights and body heights. A similar positive relationship between size and performance was found in Guatemala. The tests used in this study provided measures of visual, haptic, and kinesthetic sensory integration. Children exposed to severe early malnutrition exhibited perceptual defects as well as smaller body size. The earlier the malnutrition, the more profound the psychological retardation. The most severe retardation occurred in children admitted to the hospital under six months of age and did not improve on serial testing even after 220 days of treatment. Admitted later, children with the same socioeconomic background and the same severe malnutrition but a different time of onset did recover after prolonged rehabilitation.

Cravioto *et al.* conclude that nutritional inadequacy may interfere with both the staging and the timing of development of the brain and of behavior. Their demonstration of delayed neurointegrative development in children who have grown poorly because of malnutrition has important implications for more complex psychological functioning: "Evidence already exists that the lag in the development of certain varieties of intersensory integration has a high correlation with backwardness in learning to read. Studies of reading disability in British [and] American school children have shown that backwardness in reading is strongly associated with inadequacy in auditory-visual integration. Skill in visual-kinesthetic integration is found to be highly and significantly correlated with design copying in normal children. If it is recognized that such visual-motor control is essential in learning to write, it becomes apparent that inadequacy of intersensory organization can inter-

fere with a second primary educational skill—learning to write.

"Thus, inadequacies of intersensory development can place the child at risk of failing to establish an ordinary normal background of conditioning in his pre-school years and at the risk of failing to profit from educational experience in the school years."

In addition to the negative impact of malnutrition on the growth rates and intersensory development of children, Cravioto found a relationship between these aspects of development and infection. Eichenwald[15] has shown that certain infections in malnourished children may produce severe and prolonged hypoglycemia, a condition which can by itself cause brain damage. In addition, various biochemical defects of children with malnutrition are accentuated by infection. There is the further possibility that many infectious diseases, or the treatment of these diseases, result in damage to the nervous system that is not necessarily evident during the acute stage of the illness. Eichenwald concludes by saying: "Infection and malnutrition thus act synergistically to produce a chronically and recurrently sick child less likely to react to sensory stimuli from his already inadequate social environment."

Effects of Prenatal Nutrition

Further evidence of the interrelationship of poverty, malnutrition, illness, and the development of the child is found in a study which identifies undernutrition of poor urban mothers as the cause of the low birth weight of their offspring.[16] Since it has been found that both low birth weight and a high infant mortality rate are more common in poor families, the finding that undernutrition appears to be the cause of prenatal growth retardation is an important one. In addition to being 15 percent smaller in body weight, the infants from poor families had multiple evidences, in terms of the relative weight of such organs as the thymus, spleen, liver, etc., of prenatal undernutrition. The biggest difference found was in the relative weight of the thymus. The offspring of non-poor families had a mean thymus weight which was 104 percent of the "normal" weight, while the poor infants had a mean thymus weight of only 66 percent of "normal." While the function of the thymus[17] is not yet completely understood, there is increasing evidence that it is involved in both growth and immunological functions.

Evidence that the nutrition of the mother is one of the important variables related to the intellectual performance of the child has also been increasing. Children whose mothers received a vitamin supplement during pregnancy had significantly higher I.Q.'s at three and four years of age than did children whose mothers received placebos.[18] Erickson,[19] on the basis of a later study, also found that "when vitamin supplementation was given to pregnant and lactating women with poor nutritional environment, the offspring at four years of age had an average I.Q. score eight points greater than the average score of the children of mothers given a placebo over the

same period." Kennedy's study of prenatal nutrition on the general measures of intellectual and physical health yielded similar results.[20]

Nutritional Therapy and I.Q.

A number of other studies have been done which have also demonstrated that nutritional therapy of the child may have a beneficial effect on intellectual performance. Kugelmass *et al.*[21] demonstrated an increase in the I.Q. of both retarded and mentally normal children as a result of prolonged nutritional rehabilitation. The children, ranging in age from 2 to 10, were divided into two groups, those who were malnourished and those who were well-nourished. Each of these groups included retarded and normal children. The *malnourished* retarded children showed a gain of 10 points and the normal children one of 18 points after a period of dietary improvement. In contrast, there was relatively little change in the scores of the *well-nourished* retarded and normal children.

Harrell[22] found that in closely matched groups of presumably normal orphanage children, the double-blind daily administration of a placebo versus a 2 m.g. thiamine tablet for one year produced a superior mental response in the vitamin group. Muecher and Gruenwalt[23] demonstrated that improvement of intellectual functions can be extended even into the late teens. Students who received vitamin and mineral supplements showed a significant improvement in the performance of mental arithmetic as compared to students receiving a placebo.

Coursin[24] has shown that deficiencies in the B-complex vitamins and in vitamin C can produce abnormalities of nerve cell metabolism and function and can impair mental development. Vitamin therapy has effected improved mental functioning with such children.

Motivation and Personality Changes

In addition to the relatively direct effects of malnutrition and illness on learning, there are many more indirect effects which are difficult to isolate. It has often been observed that the malnourished child is more likely to be apathetic, irritable, and lack a long attention span. "One of the first effects of malnutrition is a reduction of the child's responsiveness to stimulation and the emergence of various degrees of apathy. Apathetic behavior in its turn can function to reduce the value of the child as a stimulus and to diminish the adults' responsiveness to him. Thus apathy can provoke apathy and so contribute to a cumulative pattern of reduced adult–child interaction. If this occurs it can have consequences for stimulation, for learning, for maturation, and for interpersonal relations, the end result being significant backwardness in the performance of later more complex learning tasks."[25]

There is an increasing amount of evidence which indicates that severe malnutrition of young infants produces significant brain damage. Mild malnutrition occurring prenatally or in the first six months of life may also produce negative effects on learning and growth. This is the beginning of a

vicious cycle of illness, slow development of intellectual functions, and inhibited growth which in turn prevent the individual from being able to protect his own children from the ravages of poverty. The *time* at which malnutrition occurs is crucial to the further development of the child. The earlier the malnutrition the more severe the effects and the more likely that they cannot be reversed. There is evidence that if malnutrition occurs after a certain age its effects on learning are reversible. Though the exact nature of the timing is yet to be worked out, it is becoming clear that the prenatal period and the first six months of life are critical.

In view of these findings, what can be done to insure that every child in America has the same chance of being "created equal"? Obviously, those most at risk are pregnant women and infants. The school would hardly seem to be the ideal setting for reaching this part of the population. Consider, however, the enormous percentage of first babies born to teen-age mothers. If the vital importance of prenatal and early infant nutrition could be conveyed to these girls, not only would they spare themselves many of the complications of pregnancy and delivery[26,27] but they would also ensure the healthy development of their babies.

In addition, there are the benefits, both short- and long-range, of providing all school children, and pre-school children when possible, with high quality breakfasts and lunches. There is evidence that malnutrition is increasing regardless of social class,[28] and not only poor children would benefit from these meals. Finally, there is the effort which as citizens and consumers we can make to insure that key staple foods are upgraded by enrichment, that the nutritional knowledge of those purchasing food and planning meals is improved, and *that food is made more readily available to those in real need.*

The most noxious of poverty's effects is malnutrition and the most important of the causes of malnutrition is poverty. Until this cycle is broken, the success of all our other efforts, such as compensatory education or remedial education, will be limited if not doomed to failure. As Margaret Mead has so eloquently put it, "Human beings have maintained their dignity in incredibly bad conditions of housing and clothing, emerged triumphant from huts and log cabins, gone from ill-shod childhood to Wall Street or the Kremlin . . . but food affects not only man's dignity but the capacity of children to reach their full potential, and the capacity of adults to act from day to day. . . . It is true that the starving adult, his efficiency enormously impaired by lack of food, may usually be brought back again to his previous state of efficiency. But this is not true of children. What they lose is lost for good . . . deprivation during prenatal and postnatal growth can never be made up."[29]

References

1. Fitzgerald: "The rich are different from you and me." Hemingway: "Yeah, they have more money."
2. C. U. Lowe. "Child Health and Public Policy." *Grover Powers Memorial Lecture.* New Haven, Conn.: Yale University Press, 1969, p. 16.

3. A. Kushlick. "Assessing the Size of the Problem of Subnormality," in J. E. Meade and A. S. Parkes (eds.), *Genetic and Environmental Factors in Human Ability.* New York: Plenum Press, 1966, p. 121.
4. M. Winick. "Malnutrition and Brain Development." *Journal of Pediatrics,* May, 1969, p. 667.
5. B. S. Platt, C. R. C. Heard, and R. J. C. Stewart. "Experimental Protein-Calorie Deficiency," in H. N. Munro and J. B. Allison (eds.). *Mammalian Protein Metabolism,* Vol. 2 New York: Academic Press, Chap. 21.
6. R. H. Barnes, A. U. Moore, I. M. Read, and W. G. Pond. "Effect of Food Deprivation on Behavioral Patterns," in N. S. Scrimshaw and J. E. Gordon (eds.), *Malnutrition, Learning, and Behavior. Cambridge,* Mass.: MIT Press, 1968, p. 168.
7. J. J. Cowley. "Time, Place, and Nutrition: Some Observations from Animal Studiess," in Scrimshaw and Gordon, *op. cit.,* p. 218.
8. M. Winick, *op. cit.*
9. M. B. Stoch and S. M. Smythe. "Undernutrition During Infancy, and Subsequent Brain Growth and Intellectual Development," in Scrimshaw and Gordon, *op. cit.,* p. 269.
10. M. Winick and P. Rosso. "Head Circumference and Cellular Growth of the Brain in Normal and Marasmic Children." *Journal of Pediatrics,* May, 1969, p. 774.
11. F. Monckeberg. "Nutrition and Mental Development." Paper presented at the Conference on Nutrition and Human Development. East Lansing, Mich., 1969.
12. Stoch and Smythe, *op. cit.*
13. R. E. Klein and O. Gilbert. "Malnutrition and Intellectual Development." Paper presented at the XI Inter-American Congress of Psychology, Mexico City, Mexico, 1967.
14. J. Cravioto, E. R. DeLicardie, and H. G. Birch. "Nutrition, Growth, and Neurointegrative Development: An Experimental and Ecologic Study." *Pediatrics Supplement,* Vol. 38, No. 2, Part II, 1966, p. 319.
15. H. F. Eichenwald and P. C. Fry. "Nutrition and Learning." *Science,* February, 1969, p. 664.
16. R. L. Naeye, M. M. Diener, W. S. Dellinger, and W. A. Blane. "Urban Poverty: Effects of Prenatal Nutrition." *Science,* November, 1969, p. 1026.
17. V. Defendi and D. Metcalf. *The Thymus.* Philadelphia: Wistar Institute Press, 1964.
18. R. F. Harrell, E. R. Woodyard, and A. J. Gates. "Influence of Vitamin Supplementation of Diets of Pregnant and Lactating Women on the Intelligence of Their Offspring." *Metabolism,* September, 1956, p. 555.
19. M. T. Erickson. "Intelligence: Prenatal and Preconception Environmental Influences." *Science,* September, 1967, p. 1210.
20. W. A. Kennedy, addendum to "Intelligence: Prenatal and Preconception Environmental Influences." *Science,* September, 1967, p. 1210.
21. I. N. Kugelmass, L. E. Poull, and E. L. Samuel. "Nutritional Improvement of Child Mentality." *American Journal of Medical Science,* November, 1944, p. 631.
22. R. F. Harrell. "Mental Response to Added Thiamine." *Journal of Nutrition,* March, 1946, p. 283.
23. H. Muecher and G. Gruenwald. "Pharmacologic Stimulation of Arithmetic Performance and Graphomotor Expansion." *Perceptual and Motor Skills.* August-December, 1962, p. 101.
24. D. B. Coursin. "Effects on Central Nervous System Function." *Nutrition Review.* March, 1965, p. 65.
25. J. Cravioto and B. Robles. "Evaluation of Adaptive and Motor Behavior During Rehabilitation from Kwashiorkor." *American Journal of Orthopsychiatry.* April, 1965, p. 449.
26. R. H. J. Hamlin. "The Prevention of

Eclampsia and Pre-Eclampsia." *Lancet.* January, 1952, p. 64.

27. T. H. Brewer. *Metabolic Toxemia of Late Pregnancy: A Disease of Malnutrition.* Springfield, Ill.: Charles C. Thomas, 1966.
28. N. Kotz. *Let Them Eat Promises.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.
29. M. Mead. "The Changing Significance of Food." *American Scientist.* March-April, 1970, p. 176.

9.2 *Mother–Child Interaction: Social Class Differences in the First Year of Life*

STEVEN R. TULKIN AND JEROME KAGAN

Reprinted from the *Proceedings of the 78th annual convention of the American Psychological Association,* 1970, **5,** 261–262, with permission of the authors and the American Psychological Association, Inc. Steven R. Tulkin is on the faculty of the State University of New York at Buffalo, and Jerome Kagan is at Harvard University.

Many observers have reported behavioral differences between children from lower and middle social classes. For one thing, lower-class children tend to have greater difficulties in "relating" to adults; for another, they seem less responsive to the verbal aspects of their environment. This study by Tulkin and Kagan sheds some light on the origins of that difference. In spite of the fact that their sample is limited as to number, race, and sex, their findings clearly suggest that middle-class mothers initiate a great deal more communication with their children and also do more reinforcing of verbal responses. This difference in child–parent interaction has obvious significance for later patterns of school-related behavior.

Investigations of social class differences in cognitive development have reported deficits in the verbal abilities of children from lower-class families (Deutsch, 1967). Hess and his colleagues have attempted to relate these deficits to differences in maternal verbal behaviors and have reported that middle-class mothers engage in more "meaningful" verbal interchanges with their preschool children than working-class mothers (Hess & Shipman, 1965). It is less clear, however, if mothers' behaviors with younger children also reflect social class differences. Descriptive reports of the experiences of young children in poor families (Pavenstedt, 1967) suggest that social class differences exist, but there has been little systematic observation of maternal behaviors with young infants.

Method

The present study reports data collected from 30 middle-class and 30 working-class Caucasian mothers. Middle class was defined by (*a*) both parents having attended college and (*b*) the father working in a professional job. Working class was defined as (*a*) either one or both parents having dropped out of high school (but neither having any college) *or* (*b*) the father working in a semiskilled or unskilled job. Each mother was observed at home on two separate days with her firstborn baby girl who was approximately 10 mo. of age. Total observation time was 4 hr.* Mothers were urged to behave naturally so that the infants would engage in their normal activities. The *O* carried a small battery-operated timer which, every 5 sec., emitted a soft tone which *O* heard through an earphone. Any of the predesignated behaviors occurring in a particular 5-sec. interval were noted on the code sheets. Although there was no class difference in the amount of infant fretting or crying, all maternal behaviors within 1 min. after an infant fret were analyzed separately to control for maternal behaviors which could be attributable to differences among the infants.

*A complete discussion of *S* selection, reliability of coding categories, and other methodological issues is given in Tulkin (1970).

Results

Environmental variables

1. The working-class children's environment provided less opportunity to experience "distinct" or "meaningful" auditory stimulation from their mothers. The infants lived in more crowded homes ($p < .001$), had more interaction with adults other than their mothers ($p < .05$), and spent more time in front of television sets ($p < .001$) than their middle-class counterparts.

2. Working-class infants also had less opportunity to explore and manipulate their environments. They had somewhat fewer toys ($p =$ approximately .10) and fewer objects such as pots and pans to play with ($p < .01$), and were more confined to the playpen or high chair ($p < .01$).

Maternal behavior

The majority of social class differences centered around the mother's verbal behavior and her attempt to "keep the

infant busy." Other aspects of mother-child interaction did not reveal social class differences. There was no class difference in the amount of time mothers spent in close proximity to their infants (within 2 ft.), although the middle-class mothers more often placed their infants in a face to face position ($p < .05$). There were also no significant differences for frequencies of kissing, holding, or tickling. Contrary to expectations, no class differences were found in the frequency of maternal prohibitions, even when the amount of time the infant was free to crawl around and explore was controlled for.

The paucity of social class differences for nonverbal variables was in sharp contrast to the dramatic differences found for the mothers' verbal behaviors. Every verbal behavior coded was more frequent among middle-class mothers ($p < .01$ for all). Middle-class mothers more often initiated vocalization, were more likely to vocalize to their infants within 5 sec. of the infants' vocalizations, and more often imitated the infants' vocalizations. Finally, middle-class mothers more frequently verbally praised their infants.

There was no social class difference in the infants' tendencies to vocalize spontaneously, suggesting that the differences in maternal vocalization were not attributable to initial differences among the infants.

Middle-class mothers more often entertained their infants ($p < .01$) by showing pictures, playing peek-a-boo, and making nonlanguage sounds, and more often gave their infants things to play with ($p < .01$).

Middle-class mothers responded to a higher percentage of the infants' frets ($p < .01$) and responded more quickly ($p < .05$). They were also more likely than working-class mothers to respond verbally when their infants fretted. Working-class mothers responded to a fretty child by giving her a bottle or a cookie, and in general made less use of themselves as agents for reducing the infants' distress.

One source of class differences in maternal behavior appeared to be the mothers' conceptions of what their infants were like. Interviews with several mothers observed in the present study suggested that working-class mothers less often believed that their infants possessed the ability to feel "adult-like" emotions or to communicate with other people. One working-class mother who constantly spoke with her daughter lamented that her friends chastized her for "talking to the kid like she was three years old." A common working-class philosophy appeared to be that only after the child began to talk was it important for the mother to speak back. Further, working-class mothers tended to feel that they could have little influence on the development of their children. They believed that an infant is born with a particular set of characteristics and that the environment exerted only minimal influence. Intervention programs, then, cannot simply focus on maternal behavior, but rather must help mothers to learn more about child development and to become more sensitive to their own children's progress.

Other differences in maternal behavior seemed more related to parental values than to different conceptions of

infancy. These behaviors, then, might legitimately be seen as reflecting "cultural" differences. Some mothers, for example, believed that children should be able to explore and discover things for themselves in an atmosphere of minimal adult-imposed structure; other mothers believed in early teaching of "right and wrong" and worried about their children becoming spoiled brats. Infants in the former group spent less time in playpens, were prohibited less, and were allowed to play with more environmental objects. The opposite was true of infants in the latter group. In each case, mothers had fairly specific ideas about the type of child behaviors they were attempting to develop and acted in accordance with this model.

Working-class mothers often felt that they could exert little influence on the development of their children, and this is probably indicative of a general sense of fatalism which develops when working-class people find that they are often powerless to effect changes in their environment. Economic pressures were also manifested by the finding that more working-class mothers worked full time, while middle-class mothers who worked were employed part time ($x^2 = 5.85$, $p = .054$). Some working-class mothers did engage in reciprocal vocalization and interactive play with their infants; but their fatigue after working and the time they had to spend caring for nieces and nephews and cooking dinner for extended families limited their availability to their infants.

References

Deutsch, M. *The Disadvantaged Child.* New York: Basic Books, 1967.

Hess, R. D., and Shipman, V. C. Early experience and the socialization of cognitive modes in children. *Child Development,* 1965, **36**, 869–886.

Pavenstedt, E. (Ed.) *The Drifters.* Boston: Little, Brown, 1967.

Tulkin, S. R. Mother-infant interaction in the first year of life: An inquiry into the influences of social class. Unpublished doctoral dissertation, Harvard University, 1970.

9.3 *Measures of Intelligence on Southwest Indian Students*

BERT P. CUNDICK

Reprinted from the *Journal of Social Psychology,* 1970, **81,** 151–156, with permission of the author and the Journal Press. Bert P. Cundick is on the psychology faculty of Brigham Young University.

It has been said that American Indian students are the most socially deprived children in our schools today. The study by Cundick explores some of the dimensions of the problem and reports the usual finding that Indian children score far below national norms on verbal measures of intelligence and that their IQ's tend to decline even further after the first year or two of school. The nonverbal IQ's of these children, however, are approximately at national norms and hold up very well. Nonverbal IQ is often taken as an indicator of a child's "true" potentiality, but the acquisition of skills such as reading and writing appears to be linked to whatever verbal intelligence tests measure. Some experimental methods designed to increase school-related cognitive abilities are discussed in the following paper by Gray and Klaus.

A. Introduction

The United States is experiencing a period of flux with respect to the treatment and consideration of minority groups. Most publicity and national interest has focused on the minority groups in our urban centers. However, the largest minority group in many of the Western States is some tribal group of American Indians. They are increasing rapidly in numbers due to a gradual lowering of their death rate and the maintenance of a high birth rate. In the past half century, they have not posed major problems for the dominant white culture because they have generally maintained a temporal and cultural separateness.

It may be that in many parts of the United States, this will continue to be the case for the immediately foreseeable future. However, in the Southwestern United States influences are at work that may well modify the situation. The Indian population is rising rapidly; vastly improved highway systems are increasing; mass communication systems are becoming important; the economic potentials of the reservation areas are being exploited; and the maintenance of separateness is becoming more difficult.

In the past five years, there has been a major influx of Indians, largely Navajo with some Utes, into the public school system in a small western town. It was the purpose of this study to do individual intellectual assessments on some of the younger children to discover how their performance would compare with national standardization samples for the instruments used and also determine their readiness for inclusion into a regular classroom. It was also decided to do some intellectual measurement on all Indian students in one elementary school to determine if the age of the students was significantly related to changes in test performance.

Prior research, dating back for many years, has generally shown Indian performance to be inferior to that of national standardization samples (1, 3, 4, 6, 7, 10, 11). The major deficits are usually on verbal type tasks. Snider has concluded that with regard to the difference between Indian and standardization samples the "greatest differences exist in the cases of the tests with the highest verbal saturation" (9, p. 42). Performance type tasks show higher means that approach, and at times exceed, the means of standardization samples (5, 6). Increasing contact with the dominant culture tends to decrease test performance differences (9), but with bilingual groups there appears to be ample evidence to suggest modifications in teaching methods and curriculum.

B. Method

1. Testing procedure and subjects

The names, birthdates, and tribal affiliations of all Indian children attending the elementary school were first obtained. The Wechsler Pre-School and Primary Scale of Intelligence (WPPSI) was administered to 27 children and the Wechsler Intelligence Scale for Children (WISC) was administered to 26 more. These children were in a special prekindergarten class, a kinder-

garten class, the first or second grades. All available Indian children in the school—72 in all—were subsequently given individual Peabody Picture Vocabulary Tests (PPVT) and the Goodenough-Harris Draw-a-Man Test (DAM). Three drawings were obtained for the DAM: a drawing of a man, a drawing of a woman, and a drawing of one's self. The self-drawing IQ was obtained by using the table for the drawing of a person of the subject's own sex.

2. *Data treatment*

Means and standard deviations were computed for each of the groups and compared with the means and standard deviations found in the manual for each of the tests. The significance of the difference between obtained means and standardization population means was then computed.

C. *Results*

A comparison between the WPPSI results of the Indian children and the manual norms appears in Table 1. The group scores significantly below the expected means on all the verbal tasks ($< .001$) and on three of the five performance tasks.

The prekindergarten children score significantly below the norm mean of 10 on all but two of the performance subtests, and these are also below the expected mean. The kindergarten children do much better as a group. Only

TABLE 1. WPPSI Means and Standard Deviations Compared with Test Norms [a]

WPPSI measure	Grade: Prekindergarten ($N = 7$) Mean	*SD*	Kindergarten ($N = 20$) Mean	*SD*	Total ($N = 27$) Mean	*SD*
Information	1.3**	.5	5.2**	1.8	4.3**	2.3
Vocabulary	2.8**	1.5	5.4**	2.6	4.8**	2.5
Arithmetic	2.3**	1.8	5.8**	2.5	5.0**	2.7
Similarities	1.8**	.4	5.0**	2.9	4.1**	2.9
Comprehension	1.7**	.5	1.2**	1.8	1.3**	1.6
Sentences	3.4**	1.1	1.2**	1.8	1.8**	1.6
Verbal IQ	50.0**	4.5	68.5**	11.6	63.9**	12.9
Animal house	7.5	1.4	8.4*	2.0	8.1*	1.9
Picture completion	2.7**	2.7	8.4*	3.2	7.2**	3.8
Mazes	6.7*	1.6	10.7	3.4	9.7	3.5
Geometric design	7.3	2.3	10.7	2.7	9.9	2.9
Block design	4.8**	3.3	9.6	2.8	8.4**	3.5
Performance IQ	71.2**	7.7	97.6	12.0	91.1**	15.7
Total IQ	56.2**	5.7	80.6**	10.9	74.6**	14.2

[a] See Wechsler, D. Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corp., 1966. pp. 18–19.

*$p < .05$.

**$p < .01$.

TABLE 2. WISC Means and Standard Deviations Compared with Test Norms [a]

WISC measure	Grade: First ($N = 10$) Mean	First SD	Second ($N = 9$) Mean	Second SD	Third ($N = 7$) Mean	Third SD	Total ($N = 26$) Mean	Total SD
Information	3.5**	1.9	3.8**	.7	4.0**	2.2	3.8**	1.5
Comprehension	3.2**	2.1	3.5**	2.6	3.7**	2.0	3.4**	2.1
Arithmetic	5.9**	2.6	6.4**	2.3	5.7**	2.2	6.0**	2.3
Similarities	3.9**	1.7	5.4**	2.3	5.9**	4.3	4.8**	2.9
Vocabulary	2.6**	2.1	3.8**	1.5	3.5**	3.6	3.2**	2.3
Digit span	5.8**	1.5	8.1	3.2	8.5	2.6	7.2**	2.8
Verbal IQ	63.1**	9.2	69.3**	11.5	66.7**	13.3	65.8**	11.1
Picture completion	11.2	2.7	10.6	3.1	8.9	3.8	10.3	3.2
Picture arrangement	7.3*	2.9	8.1	2.6	8.6	3.2	8.0**	2.7
Block design	8.8	2.0	9.8	2.0	9.6	4.1	9.3	2.8
Object assembly	9.1	1.7	10.6	1.8	9.3	3.5	9.7	2.4
Coding	8.8	2.1	10.3	2.5	10.1	3.7	9.4	2.8
Mazes	9.3	1.4	9.0	1.2	10.3	1.6	9.4	1.3
Performance IQ	94.1	8.9	98.6	6.7	95.0	20.7	95.9	12.0
Total IQ	75.5	9.2	81.7	8.6	78.3	16.1	78.1**	11.0

[a] See Wechsler, D. Manual for the Wechsler Intelligence Scale for Children. New York: Psychological Corp., 1949. p. 15.

*$p < .05$.

**$p < .01$.

two of the performance subtest scores are significantly below the norm mean ($< .05$), and the performance IQ is not significantly different from the norm IQ of 100. Although the verbal subtest scores and verbal IQ are much improved over their prekindergarten counterparts, all of them are much below the norm means ($< .001$). The WISC results are found in Table 2. Performance is similar to that on the WPPSI. All verbal tasks for the group show significantly lower than expected mean scores. However, only one performance task, the picture arrangement, is significantly lower than the expected mean. The WISC IQ scores increase for the group until the second grade.

The data for the DAM and the PPVT are found in Table 3. It should be noted that although the means for each of the drawings are significantly below the expected mean for the group, the grade means are about at the expected mean of 100 by the third grade except for the third drawing. This is a self-portrait which has not been adequately standardized according to Harris.

The PPVT results present a very different picture. The means increase until the second grade and then drop off. At all grade levels the means are significantly below 100.

D. Discussion

The relatively normal performance IQ's for this group on the Wechsler Scales

TABLE 3. Goodenough-Harris DAM and PPVT Means and Standard Deviations Compared with Test Norms [a, b]

Grade	Test measure							
	DAM 1		DAM 2		DAM 3		PPVT	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Kindergarten ($N = 19$)	89.8**	10.6	91.9	13.5	87.4**	11.7	53.4**	20.9
First ($N = 17$)	90.9*	10.6	93.9	14.8	94.9	13.4	62.2**	13.7
Second ($N = 16$)	97.3	16.1	94.7	11.5	94.8	12.7	69.1**	9.4
Third ($N = 7$)	104.3	19.5	97.6	13.9	98.4	15.8	66.4**	11.0
Fourth ($N = 9$)	102.4	18.7	102.8	20.5	97.2	18.8	67.6**	14.1
Fifth ($N = 13$)	98.8	12.8	96.1	12.9	93.9	11.6	58.2**	14.0
Sixth ($N = 10$)	102.6	17.1	99.3	14.4	95.1	15.8	61.8**	10.6
Total ($N = 91$)	96.3	14.9	95.6	14.2	93.6	13.7	61.4**	15.6

[a] See Harris, D. B. Children's Drawings as Measures of Intellectual Maturity. New York: Harcourt, Brace & World, 1963. p. 89.

[b] See Dunn, L. M. Expanded Manual for the Peabody Picture Vocabulary Test. Minneapolis, Minn.: American Guidance Service, 1965. p. 10.

*$p < .05$.

**$p < .01$.

and the DAMs suggest that performance tasks can be used with Indian children populations in the Southwest United States and will provide normative data that will roughly approximate standardization samples, provided the children have had at least one year of schooling.

Verbal tasks will not yield IQ distributions that are normal even with older children. Although the data in this study are cross-sectional and do not yield a picture of IQ growth in individual children, a number of implications can be drawn from the data. The failure to increase in verbal IQ after the second grade reveals a problem of significance. It appears that these children may fall further behind in verbal skills as they remain in the school system. The drop for this group is similar to the IQ drop between the third and sixth grades reported by Clark for ghetto children in New York (2, p. 121); although the reasons for the drop may be quite different, there may be parallels.

Rosenthal's finding that the IQ's of Mexican-American students, particularly those who "appeared most Mexican," were most affected by a positive prediction about their intellectual growth during the year may have relevance here (8). Both the Clark and Rosenthal work suggest the presence of a "self-fulfilling" prophecy where children tend to perform as they are expected to perform once they have internalized the expectancies of those around them.

It may also be that the emphases upon primary educational skills in arithmetic and reading in the first years of elementary school change at this time

and language skills become more important as instruction moves more rapidly.

Summary

IQ scores on the WPPSI, WISC, PPVT, and Harris-Goodenough Draw-a-Man for Indian children attending the same public elementary school were obtained. Performance tests yielded nearly normal IQ's after one year of schooling. Verbal IQ's in this school did not significantly increase after the second grade. Possible reasons for the failure to increase may be a result of a combination of increased instructional speed and a greater emphasis on language. However, the failure to increase may be a result of a "self-fulfilling" prophecy.

References

1. Altus, W. D. A note on group differences in intelligence and the type of test employed. *J. Consult. Psychol.,* 1948, **12**, 194–195.
2. Clark, K. B. *Dark Ghetto.* New York: Harper & Row, 1965.
3. Fitzgerald, J. A., and Ludeman, W. W. The intelligence of Indian children. *J. Comp. Psychol.,* 1926, **6**, 319–328.
4. Garth, T., and Smith, O. The performance of full-blooded Indians on language and non-language intelligence tests. *J. Abn. & Soc. Psychol.,* 1937, **34**, 376–381.
5. Havighurst, R. J., and Hilkevitch, R. R. The intelligence of Indian children as measured by a performance scale. *J. Abn. & Soc. Psychol.,* 1944, **39**, 419–433.
6. Howell, R. J., Evans, L., and Downing, L. N. A comparison of test scores for the 16-17 year age group of Navajo Indians with standardization norms for the Wechsler Adult Intelligence Scale (Arizona and New Mexico). *J. Soc. Psychol.,* 1958, **47**, 355–359.
7. Joseph, A. *The Desert People.* Chicago, Ill:. Univ. Chicago Press, 1949.
8. Rosenthal, R., and Jacobson, L. *Pygmalion in the Classroom.* New York: Holt, Rinehart & Winston, 1968.
9. Snider, J. G. Achievement test performance of acculturated Indian children. *J. Educ. Res.,* 1961, **7**, 39–41.
10. Snider, J. G., and Coladarci, A. P. Intelligence test performance of acculturated Indian children. *Calif. J. Educ. Res.,* 1960, **11**, 34–36.
11. Turner, G. H., and Penfield, D. J. The scholastic aptitude of Indian children of the Caradoc Reserve. *Can. J. Psychol.,* 1952, **6**, 31–44.

9.4 *The Early Training Project: A Seventh Year Report*

SUSAN W. GRAY AND RUPERT A. KLAUS

Reprinted from *Child Development*, 1970, **41**, 909–924, with permission of the authors and the Society for Research in Child Development, Inc.; Copyright © 1970 by The Society for Research in Child Development, Inc. Susan W. Gray is a developmental psychologist who has played a leading role nationally in the school psychologist profession. Rupert A. Klaus is a school psychologist who serves as a consultant to the Murfreesboro City Schools in Tennessee and also as the director of the Early Training Project described in this article. Both authors are on the staff of George Peabody School for Teachers in Nashville, Tennessee.

The fact that children from poverty homes tend to have more than their fair share of problems in learning has been related to their preschool environment. Not only do they have fewer relationships with adults, as the study by Tulkin and Kagan suggests, but they also are not encouraged to develop the attitudes and skills that are basic to cognitive development, particularly in verbal areas, as the study by Cundick of Southwest Indian children showed. The question of how these deficiencies can be remedied has led psychologists to experiment with various kinds of intervention programs. The Early Training Program at George Peabody College for Teachers in Nashville, Tennessee, is in many ways a model experimental program in its experimental design and in the employment of a wide variety of intervention modalities.

The Early Training Project has been a field research study concerned with the development and testing over time of procedures for improving the educability of young children from low income homes. The rationale, the general design and methodology, and findings through the second year of schooling have been reported in some detail in *The Early Training Project for Disadvantaged Children, a Report after Five Years,* by Klaus and Gray (1968). A briefer report, up to school entrance, is given in Gray and Klaus (1965). The purpose of this report is to present the findings at the end of the fourth grade, three years after all experimental intervention had ceased.

The major concern of the Early Training Project was to study whether it was possible to offset the progressive retardation observed in the public schooling careers of children living in deprived circumstances. In addition, the writers undertook to study the spillover effect upon other children in the community and upon other family members.

The general research strategy was one of attempting to design a research "package" consisting of variables which —on the basis of research upon social class, cognitive development, and motivation—might be assumed to be relevant to the school retardation which is observed in deprived groups and which at the same time might be subject to the effects of manipulation. Because this was a problem with major social implications, we also tried to design a general treatment approach which it would be feasible to repeat on a large scale, in the event that the procedures proved successful.

Subjects were 88 children born in 1958. Sixty-one of these lived in a city of 25,000 in the upper South. The remaining 27, who served as a distal control group, resided in a similar city 65 miles away. The children were all Negro. When we initiated the study the schools of the city were still segregated; we chose to work with Negro children because in this particular setting we had reason to believe that our chances of success were greater with this group.

The children were selected on the basis of parent's occupation, parent's education, income, and housing conditions. At the beginning of the study incomes were considerably below the approximate $3,000 used as the poverty line for a family of four. Occupations were either unskilled or semiskilled; the educational level was eighth grade or below; housing conditions were poor. The median number of children per family at the beginning of the study was five; in about one-third of the homes there was no father present.

From the 61 children in the first city three groups were constituted by random assignment. The first group (T1) attended, over a period of three summers, a ten-week preschool designed to offset the deficits usually observed in the performance of children from disadvantaged homes. In addition, this group had three years of weekly meetings with a specially trained home visitor during those months in which the preschool was not in session. The second group (T2) had a similar treatment, except that it began a year later; the children received two summers of the special preschool and two years of home visits. The third group (T3) be-

came the local control group, which received all tests but no intervention treatment. The fourth group (T4), the distal control group, was added to the design because of the somewhat ghetto-type concentration of Negroes in the first city. The local and distal control groups also made possible the study of spillover effects upon children and parents living in proximity to the experimental children. The general layout of the experimental design is given in Table 1. By reading down the columns, one may see the particular treatment and testing sequence followed for each of the four groups. Periodic testing is continuing for the children through elementary school.

The Intervention Program

The overall rationale for the intervention program grew out of the literature on child-rearing patterns in different social classes, plus the writers' own observations in low income homes. On

TABLE 1. Layout of General Research Design

Treatments	T^1 Three Summer Schools	T^2 Two Summer Schools	T^3 Local Controls	T^4 Distal Controls
First Winter 1961–62	(Criterion development, curriculum planning, general tooling up)			
First Summer 1962	Pre-test Summer school Post-test	Pre-test Post-test	Pre-test Post-test	Pre-test Post-test
Second Winter 1962–63	Home visitor contacts			
Second Summer 1963	Pre-test Summer school Post-test	Pre-test Summer school Post-test	Pre-test Post-test	Pre-test Post-test
Third Winter 1963–64	Home visitor contacts	Home visitor contacts		
Third Summer 1964	Pre-test Summer school Post-test	Pre-test Summer school Post-test	Pre-test Post-test	Pre-test Post-test
Fourth Winter 1964–65	Home visitor contacts	Home visitor contacts		
Fourth Summer 1965	Follow-up tests	Follow-up tests	Follow-up tests	Follow-up tests
Fifth Summer 1966	Follow-up tests	Follow-up tests	Follow-up tests	Follow-up tests
Seventh Summer 1968	Follow-up tests	Follow-up tests	Follow-up tests	Follow-up tests

the basis of this study, the intervention program for children was organized around two broad classes of variables: attitudes relating the achievement, and aptitudes relating to achievement. Under attitudes we were particularly interested in achievement motivation, especially as it concerns school-type activities, in persistence, in ability to delay gratification; and in general interest in typical school materials, such as books, crayons, puzzles, and the like. We were also concerned with the parent's attitude toward achievement, particularly in their aspirations for their children, especially as they related to schooling.

In the broad class of aptitude variables relating to achievement we were particularly interested in perceptual and cognitive development and in language. Children from low income homes have been shown to have deficits in these areas, all of which appear closely related to school success in the primary grades.

In the summer months, for 10 weeks the children met in assembled groups. Each of the two experimental groups had a head teacher, who was an experienced Negro first grade teacher. There were in addition three or four teaching assistants. These assistants were divided about equally as to race and sex.

The work with the parents in the project was carried on largely through a home visitor program in which a specially trained preschool teacher made weekly visits to each mother and child. Both the home program and the school program are described in considerable detail in *Before first grade* (Gray, Klaus, Miller, and Forrester, 1966) and in Klaus and Gray (1968).

Prior to and after each summer session children in all four groups were tested on several instruments. From the first summer certain standardized tests of intelligence and language were used, along with a number of less formal instruments. At the end of first grade, achievement tests were added. This testing schedule is shown in Table 1. In general the .05 level of significance was used.

Results

The detailed results of the testing program through May, 1966, the end of the second grade for the children, are given in Klaus and Gray (1968). This paper gives the results as they relate to the spring and summer testings of 1968 with some additional information on performance of younger siblings. The same kinds of analyses were used for the 1968 data as were used in the earlier paper.

In 1968 the following tests were administered to all children still residing in middle Tennessee: the Binet, the Peabody Picture Vocabulary Test, and the Metropolitan Achievement Test. The analyses here reported are based only upon those children available for testing with the exception of one child in the distal control group.

The Binet scores are given in Table 2, and are portrayed graphically in Figure 1. A Lindquist (1953) Type 1 analysis of the results of 1962–1968, in terms of IQ, gave a significant F of 4.45 for the four groups, and F of 16.81 for repeated measures, and F for interac-

TABLE 2. Mean Stanford–Binet MA and IQ Scores for the Four Treatment Groups at Each Administration

Date of Administration	T1 (N = 19)		T2 (N = 19)		T3 (N = 18)		T4 (N = 23)	
	MA (mo.)	IQ	MA (mo.)	IQ	MA (mo.)	IQ	MA (mo.)	IQ
May 1962	40.7	87.6	43.8	92.5	40.3	85.4	40.3	86.7
Aug. 1962	50.7	102.0	46.9	92.3	44.3	88.2	43.4	87.4
May 1963	55.6	96.4	56.0	94.8	53.2	89.6	50.4	86.7
Aug. 1963	59.3	97.1	60.6	97.5	55.0	87.6	52.3	84.7
Aug. 1964	68.0	95.8	71.6	96.6	62.3	82.9	59.4	80.2
Aug. 1965	83.8	98.1	86.3	99.7	79.4	91.4	77.0	89.0
June 1966	88.7	91.2	93.4	96.0	86.8	87.9	82.9	84.6
July 1968	106.0	86.7	111.4	90.2	104.7	84.9	96.2	77.7

tion of groups over time of 3.51. All of these were significant at the .01 level or beyond. Next an analysis was made by the use of orthogonal comparisons. These are given in Table 3.[1]

The scores across the ten administrations of the Peabody Picture Vocabulary Test are given in Table 4 in MA and IQ form. A Lindquist (1953) Type 1 analysis of variance was performed for the MA scores. F for groups was 5.16, indicating a significant effect of the experimental treatment upon the children's performance. F for repeated testings was 376.73, an effect that would be clearly expected when MA scores were used. These were selected in preference to IQ scores on this particular test since the IQ scores appear to lack discrimination at certain levels. The interaction between groups and time was nonsignificant. Orthogonals were next used. Here was found that T1 + T2 was significantly greater than T3 + T4 up until 1968, in which year differences were not significant. As may be seen from Table 4, differences in mean scores were still apparent.[2] Heterogeneity had increased over time, however, so that differences were no longer significant. In no analysis at any point of time was either experimental group significantly superior to the other. Nor did either control group show itself to be significantly superior to the other one.

The results for the Metropolitan Achievement Test are given in Table 5. A Lindquist (1953) Type 1 analysis was performed on each subtest, and orthogonal comparisons made. In the interest of brevity a table of orthogonal comparisons is not given. In 1965, at the end of first grade, the experimental children were significantly superior on three of the four tests used at that time: word knowledge, word discrimi-

[1]Table 3 is omitted in the interest of brevity. It presents data showing that the two experimental groups remain significantly superior to the control groups and suggests that the distal control group showed more decline in IQ than the local control group.

[2]Table 4 has been omitted in the interest of brevity.

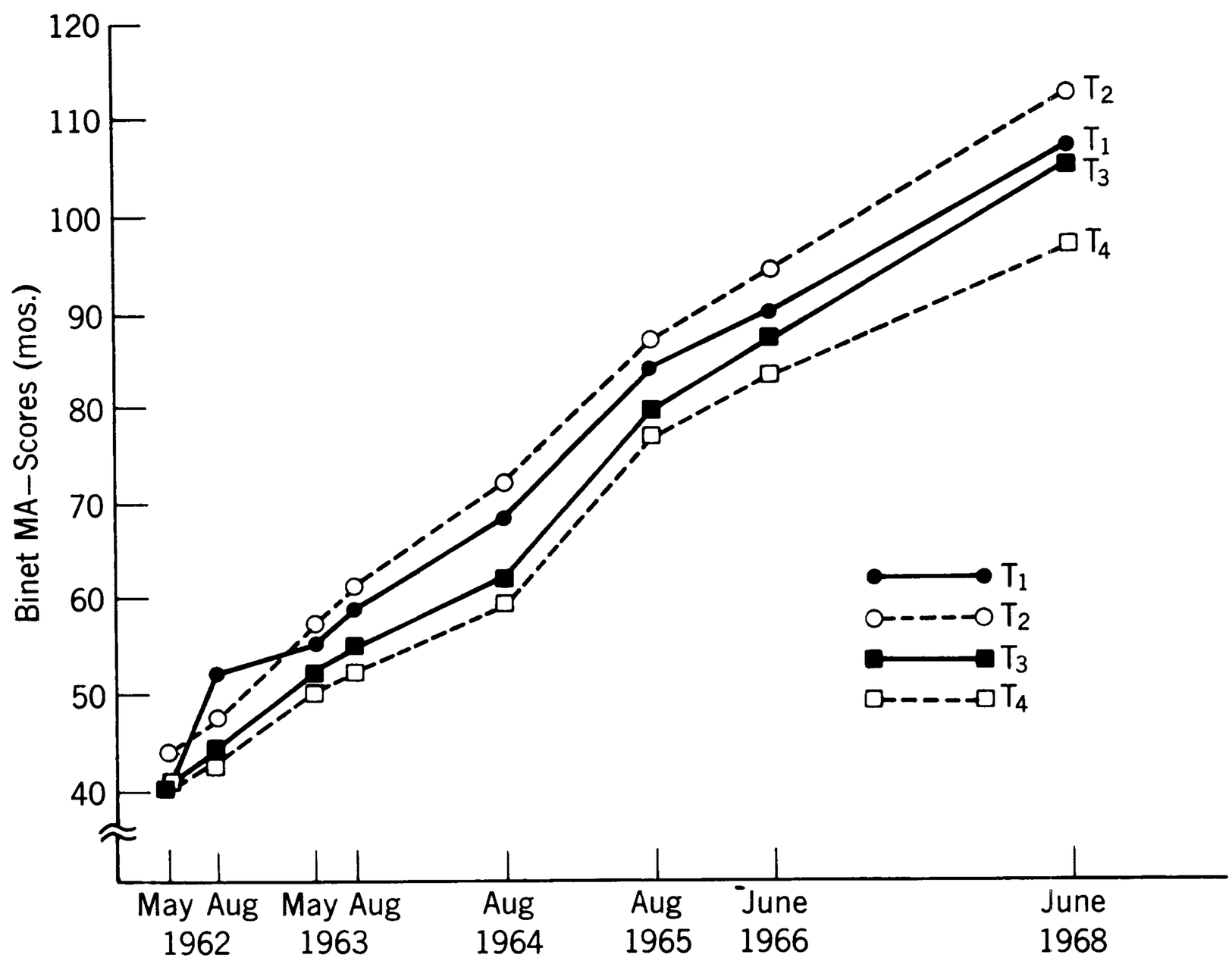


FIGURE 1. Mental ages for experimental and control groups on the Stanford-Binet.

nation, and reading. For arithmetic computation scores, *F* was less than 1.00. The local controls were also somewhat superior to the distal controls on these tests, an indication possibly of horizontal diffusion or, either in interaction or independently, a somewhat better instructional program. In 1966 five subtests were given. This time only two were significant, word knowledge and reading. On the other three tests, however, the *F*'s ranged from 2.69 to 2.84, suggesting probabilities at about the .10 level. In neither year was T1 significantly superior to T2. The highest *F* was 1.16, where $F_{.95}$ is 3.97. In the comparisons of T3 and T4, T3 was superior to T4 on reading and arithmetic computation. On word knowledge, word discrimination, and spelling the *F*'s ranged from 3.19 to 3.85, suggesting probabilities beyond the .10 level ($F_{.90} = 2.77$). At the end of the fourth year no significant effects were found with the single exception of reading, on which T3 was superior to T4. There is some suggestion of residual effect since in six of the seven possible comparisons of experimental and controls, the experimentals were superior. Also on all seven possible comparisons the local control group was superior to the distal control group.

The Binet was administered in all four groups to those younger siblings who were of testable age. This was first done in 1964 and again in 1966. Since the 1966 findings have not been

TABLE 5. Metropolitan Achievement Test Grade Equivalent Mean Scores for the Various Subtests for the Three Administrations

Subtest and Year	T1	T2	T3	T4
Word knowledge:				
1965	1.69	1.73	1.79	1.37
1966	2.32	2.47	2.29	1.98
1968	3.58	3.90	3.54	3.27
Word discrimination:				
1965	1.68	1.81	1.82	1.37
1966	2.64	2.73	2.65	2.20
1968	3.73	3.95	3.76	3.47
Reading:				
1965	1.72	1.82	1.84	1.46
1966	2.52	2.75	2.56	2.11
1968	3.52	3.89	3.72	3.10
Arithmetic computation:				
1965	1.52	1.62	1.54	1.43
1966	2.41	2.55	2.49	2.05
1968	3.92	4.07	4.06	3.79
Spelling:				
1966	2.42	2.85	2.60	1.99
1968	4.26	4.69	4.24	3.67
Language:				
1968	3.52	4.00	3.63	3.17
Arithmetic problem-solving and concepts:				
1968	3.31	3.54	3.75	3.26

previously reported they are presented here in Table 6. In 1964, 57 children were tested. Fifty of these same children were tested again in 1966, along with 43 additional siblings who were too young to test in 1964.

An analysis of co-variance was performed on these scores, with the IQ's at first testing of the target-age children used as the covariable. Also, where there were two younger siblings in the same family, one was dropped, so that the analysis was based on 87 children. Separate analyses were also performed for the 1964 and the 1966 results of all children who were retested. In addition, an analysis was performed on the 1966 results for those children who were being tested for the first time.

On all younger siblings tested in 1966 the F between groups was not significant at the .05 level ($F = 3.97$). It was significant beyond the .10 level, and therefore we made further analyses. Orthogonal comparisons were used, with the hypotheses shown in Table 7.[3] This is the same general ap-

[3] Table 7 has been omitted in the interest of brevity.

TABLE 6. Initial Binet Scores of Treatment Group Children and Younger Siblings in Two Testings

Testing	Groups	Mean Scores (First Testing, 1962) for Treatment Group Children with Younger Siblings			Mean Scores for Younger Siblings		
		N	CA	IQ	N	CA	IQ
1964 testing of younger siblings born in 1959 and 1960	T1	12	47	82	13	54	82
	T2	16	46	89	21	53	83
	T3	7	50	84	9	54	71
	T4	12	48	88	14	62	74
1966 retesting of younger siblings initially tested in 1964	T1	12	47	82	13	78	85
	T2	14	46	92	19	76	85
	T3	5	46	82	7	76	78
	T4	11	48	86	13	77	75
1966 testing of younger siblings born in 1961 and 1962	T1	10	44	87	11	58	84
	T2	9	47	91	10	52	87
	T3	7	48	83	9	56	76
	T4	12	47	88	15	55	84
1966 testing of all younger siblings	T1	15	50	84	24	69	84
	T2	17	46	91	29	68	86
	T3	8	47	84	16	65	77
	T4	15	47	86	28	63	80

proach as used with the target children. All orthogonal comparisons showed significant differences for the testing of all younger siblings in 1966: the combined experimental group siblings were superior to the combined control group siblings; the T1 siblings were superior to the T2 siblings; and the T3 siblings were superior to the T4 siblings. When the children who were tested for the first time are separated out, it is clear, both in the 1966 and the 1964 data, that most of the variance was being carried by younger siblings closer in age to the target-age children. There are some interesting implications of these general results on younger siblings which will be examined in more detail in the discussion section.

Discussion

The results on the one test of intelligence which was used consistently from the initiation of the program in 1962 until the testing at the end of the fourth grade, in 1968, are very much in line with what might be expected. For this was an intervention program that used a broad gauge approach and which was relatively successful in terms of improving the educability of young children from low income homes. Intervention caused a rise in intelligence which was fairly sharp at first, then leveled off, and finally began to show decline once intervention ceased. The control groups on the other hand tended to show a slight but consistent decline with the

single exception of a jump between entrance into public school and the end of first grade. Differences between experimentals and controls on Binet IQ were still significant at the end of the third year after intervention ceased. All four groups have shown a decline in IQ after the first grade but the decline, as shown in Figure 1, tended to be relatively parallel. Perhaps the remarkable thing is, with the relatively small amount of impact over time that differences should still be significant. After all, the child experienced only five mornings of school a week for ten weeks for two or three summers, plus weekly home visits during the only other nine months for two or three years. This suggests that the impact was not lost. It was not sufficient, however, to offset the massive effects of a low income home in which the child had lived since birth onward.

The results on the PPVT showed a pattern that is not dissimilar. There was a rise during intervention, including the first grade, then a leveling-off and a slight decline. Here, however, difference between groups, although consistent were no longer significant.

The importance of the school situation for the maintenance or loss of a gain should be weighed. The children for the most part remained in schools in which the entire population was Negro. Eight of the local children at the end of first grade did enroll in schools that had previously been all white. Four more changed during the next two years. None of the distal children attended schools with white children. Since in this area, as in many places, race tends to be confounded with social class, the children in the study did not in general have the advantage of classmates with relatively high expectancies. There is some evidence that in both of the all-Negro schools the general teaching-learning situation, although fair, was less adequate than in the schools that have formerly been all white. This, plus the continuing effect of the home situation and the immediate community, took its toll. There are some data on achievement test scores to be presented later which suggest the impact of the two all-Negro schools which most of the children attended.

On the one achievement battery administered from first to fourth grade, the Metropolitan Achievement Test (Table 5), significant differences did not appear in 1968 on any of the subtests with sole exception of the reading score, in which the local control group was superior to the distant control group. The experimentals had been superior to the controls on three tests in 1965 and on two tests in 1966. One might interpret this as showing that the intervention program did have measurable effects upon test performance at the end of first grade, but that by the end of fourth grade, the school program had failed to sustain at any substantial level the initial superiority. Although disappointing, this is perhaps not surprising in a test battery so dependent upon specific school instruction.

An interesting sidelight is thrown on this matter by looking at the performance on the Metropolitan Achievement Test of the eight children from the

local school who at the end of first grade enrolled in previously all-white schools. An attempt was made, on the basis of first grade achievement tests and home ratings of educational aspirations, to match these eight children with eight who remained in the Negro school. Admittedly, this is a chancy business, and one which should not be taken too seriously. Table 8 presents the gains in grade equivalents on the Metropolitan Achievement Tests from the end of first grade to the end of fourth grade. On the four subtests common to both grade levels the picture is a clear one of more gain in the children who changed schools, varying from .8 to 1.4 years' greater gain. These data did not seem appropriate for subjection to statistical analysis. They do suggest, however, the fairly obvious: that performance on achievement tests is directly related to school experience. The children who changed schools have made approximately "normal" gain for their three years; the children who did not change have gained two years or less during the three years from first through fourth grade.

The results on the younger siblings are to the writers among the most interesting findings of the study. We have termed the process by which such results are achieved and the product of that process as vertical diffusion, to suggest that this is a spread of effect down the family from the mother and possibly the target-age child to a younger child. In this study the effects of the older sibling and the mother upon the younger child were confounded. Some research currently being carried on under the direction of one of the writers has made possible the separation of the influence of mother and older siblings. Results so far indicate that most of the effect is coming from the mother. It is plausible to assume that the role of the mother was the more influential since considerable effort was expended by the home

TABLE 8. Mean Gains on the Mat Over a 3-Year Period for 8 ETP Children in Integrated Schools and Matches in Negro Schools

	Mean Gains 1965–68			
	Word Knowl.	Word Disc.	Read.	Arith.
ETP Ss in integrated schools beginning fall 1965	3.1	2.8	2.7	2.9
ETP Ss in Negro schools matched to the first group on spring 1965 MAT and on verbal rating by home visitor	1.7	2.0	1.6	1.7
Difference	1.4	.8	1.1	1.2

visitor over a period of three years with the first experimental group and over two years with the second experimental group. The emphasis of the home intervention was on making the mother a more effective teacher, or more generally, an effective educational change agent for her target-age child. Also worthy of note is the finding that vertical diffusion appeared more clearly in the younger siblings born in 1959 and 1960, who were within one to two and a half years in age to the older siblings. The siblings born in 1961 and 1962, when pulled out for separate analysis, did not show an effect which approached statistical significance. Vertical diffusion also appeared more operative in the first than in the second experimental group. A plausible explanation is that intervention lasted a year longer with the first group and began a year earlier. There is also in the data some suggestion of a process we have examined in more detail elsewhere (Klaus and Gray, 1968), one that may be termed horizontal diffusion, the spread of effect from one family to another. This we have in general analyzed by comparing the local and distal control groups. Here we found that the younger siblings in the local control group showed themselves to be superior to the distal control group.

To the extent that the findings on vertical diffusion have generality, they seem to point to the efficacy of a powerful process in the homes, presumably mediated by the parent, which may serve to improve the educability of young children. Before a second conclusion is reached by the reader, however, to the effect that "parent education" is the answer, we would like to point out that our procedure was clearly parent education with a difference. It was conducted in the home; it was done by skilled preschool teachers with some experience in working in the homes; it was highly concrete and specific to a given mother's life situation; it was continuous over a long period of time. Indeed, parent education probably is the answer, but in low income homes a very different kind of parent education from that usually provided may be needed.

Seven years after the Early Training Project began, in 1969, intervention programs for young children from low income homes are nationwide. These programs differ tremendously in the length and timing of the intervention, in the objectives and consistency with which they are followed, in the degree of specificity of the program, and in the length and extent of follow-up study of the sample.

It is hardly surprising, with the wilc heterogeneity of such programs, tha nationwide assessment of programs such as the Westinghouse Survey o Project Head Start (1969), would finc relatively small evidence of positive effects upon the child's achievemen and personal adequacy. Leaving aside all the problems of measuring personai adequacy and even achievement in young children, such lack of results is only to be expected in situations where the bad or inappropriate so cancels out the good that little positive effect can be found, especially if the evaluation is somewhat premature.

At this point in time it seems appropriate to look more closely at those programs which have clearly followed an adequate research design, specified and carefully monitored their treatments, and conducted adequate follow-up study of the sample. Such programs are relatively few in number, for their history is short.

In the Early Training Project we have been more fortunate than most. The study was initiated nearly four years before the tidal wave of interests in such early intervention that came about through such nationwide programs as Project Head Start and Title I and III of the Elementary and Secondary Education Act. We have worked in a setting in which we have been free from administrative pressures either to change our procedures or to make premature conclusions from our data. The two communities in which families live have had little outward mobility; even at the end of seven years attrition is only a minor problem. For these reasons we believe that data collected over seven years with our four groups of children do shed some light upon the problem of progressive retardation and the possibility that it can be offset.

Our answer as to whether such retardation can be offset is one of cautious optimism. The effects of our intervention program are clearly evidenced through the second year of public schooling, one year after intervention ceased. There is still an effect, most apparent in the Binet, after two more years of nonintervention. Our data on horizontal and vertical diffusion, especially the latter, gives us some hope that intervention programs can have a lasting effect that goes beyond the children that were the target of that intervention program.

Still, it is clear from our data, with a parallel decline across the four groups in the second through fourth grades, that an intervention program before school entrance, such as ours, cannot carry the entire burden of offsetting progressive retardation. By some standards the Early Training Project might be seen as one of relatively massive intervention. And yet a colleague of ours (Miller, 1969) has estimated that in the years prior to school entrance the maximum amount of time that the children in the project could have spent with the Early Training Project staff was approximately 600 hours, less than two percent of their waking hours from birth to six years. Perhaps the remarkable thing is that the effect lasted as well and as long as it did. In a similar vein, we have estimated the amount of these contracts which was in the home as a maximum of 110 hours, are about 0.3 percent of the waking hours of the child from birth to six years. Surely it would be foolish not to realize that, without massive changes in the life situation of the child, home circumstances will continue to have their adversive effect upon the child's performance.

In 1968 the authors wrote:

"The most effective intervention programs for preschool children that could possibly be conceived cannot be considered a form of innoculation whereby the child forever after is immune to the

effects of a low income home and of a school inappropriate to his needs. Certainly, the evidence on human performance is overwhelming in indicating that such performance results from the continual interaction of the organism with its environment. Intervention programs, well conceived and executed, may be expected to make some relatively lasting changes. Such programs, however, cannot be expected to carry the whole burden of providing adequate schooling for children from deprived circumstances; they can provide only a basis for future progress in schools and homes that can build upon that early intervention."

In 1969 we see no reason to alter this statement. Our seventh year results only serve to underscore its truth.

Summary

This is a report at the end of fourth grade of a preschool intervention project for children from low income homes. Its purpose was to investigate whether one could offset progressive retardation in elementary school. Special experiences provided for the 44 experimental children were based upon variables associated with attitudes and aptitudes conducive to school achievement. Intensive work was done for three summers; in the remaining months there were weekly home visits. Over the years the experimental children remained significantly superior to control children on intelligence tests. On measures of language and achievement trends still remained, but differences were no longer significant by the end of fourth grade. There is a slight but parallel decline across groups. Evidence is presented on younger siblings.

References

Gray, S. W., and Klaus, R. A. An experimental preschool program for culturally deprived children. *Child Development,* 1965, **36**,887–898.

Gray, S. W., and Klaus, R. A., Miller, J. O., and Forrester, B. J. *Before First Grade.* New York: Teachers College Press, Columbia University, 1966.

Klaus, R. A., and Gray, S. W. The early training project for disadvantaged children: A report after five years. *Monograph of the Society for Research in Child Development,* 1968, **33**, (4, Serial No. 120).

Lindquist, E. F. *The Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin, 1953.

Westinghouse Learning Corporation. *The Impact of Head Start: An Evaluation of the Head Start Experience on Children's Cognitive and Affective Development.* Westinghouse Learning Corporation, Ohio University, 1969.

9.5 *A Comparison of the Problems of Certain Anglo- and Latin-American Junior High School Students*

ROBERT PAUL WITHERSPOON

Reprinted from the *Journal of Educational Research,* 1960, **53,** 395–399, with permission of the author and Dembar Educational Research Services, Inc. Robert Paul Witherspoon teaches psychology and is the director of the testing center at Arkansas State Teachers College in Conway.

Spanish-speaking Americans have special problems in our schools. Those that are from the lower social classes display the kind of behavior patterns that are characteristic of socially deprived children of other cultural groups: a low verbal IQ that declines over time, coupled with difficulties in acquiring the basic skills taught by the curriculum. These problems are further complicated by bilingualism. This combination of cultural background and unfavorable school experiences further leads to or is accompanied by the development of attitudes and values which are in many ways different from those of Anglo students at the same socioeconomic level. In the present paper, Witherspoon presents the results of a survey of adjustment problems reported by adolescents of the two cultural groups. It is interesting to note that although the Latin-American students have more school-related problems, their attitudes toward teachers and toward the school in general are more positive than those of the Anglo students. This finding also appears in other studies comparing the attitudes of middle-class and lower-class students, in that lower-class students express more appreciation for school in general and have more positive attitudes toward teachers.

Several states have a large Latin-American population. One function of the school is to help its students to adjust better to their problems. The question arises whether the two major racial groups, i.e., Anglo-Americans and Latin-Americans, who attend non-segregated schools, have the same problems, or do they differ in type and acuteness.

For this study, a total of 309 students filled out the Junior High School form of the Mooney Problem Check List: sixty-three Anglo-American girls, ninety-seven Latin-American girls, seventy-one Anglo-American boys, and seventy-eight Latin American boys, in grades seven and eight of a junior high school in San Antonio, Texas. The school is located in a combination business-residential section of town and has about an equal number of Latin- and Anglo-Americans. The neighborhood as a whole is composed of a somewhat below average socioeconomic stratum.

The Mooney Problem Check List, Junior High School Form, 1950 Revision, consists of 210 problems divided into seven areas with thirty items in each area. The seven areas are: (1) Health and Physical Development (HPD), (2) School (S), (3) Home and Family (HF), (4) Money, Work, the Future (MWF), (5) Boy and Girl Relationship (BG), (6) Relations to People in General (PG), and (7) Self-Centered Concerns (SC).

The students were instructed to underline the problems troubling them, and then to circle the problems troubling them most. The usual precautions were taken to get valid responses.

The two groups are compared under four groupings, namely, Latin girls, Anglo girls, Latin boys, and Anglo boys. Only those problems marked by at least 20 percent of one of the four groups were considered in this study; and only those showing at least a 10 percent difference will be discussed. The problems of most concern are shown if at least 10 percent of one of the groups marked it.

Out of eighteen school problems, the Latin-Americans have a higher percentage in twelve. Five of the problems are common to the boys and girls. Sixteen percent more of the Latin boys and 20 percent more of the Latin girls are afraid of failing in school work than are the Anglos. Both of the two races have trouble with arithmetic, but a higher percentage of the Latins have trouble here. Feeling that they are not smart enough is indicated by 35.9 percent of the Latin boys and 41.1 percent of the Latin girls, as compared to only 11.2 percent of the Anglo boys and 11.1 percent of the Anglo girls. Also, more Latin boys and girls are afraid to speak up in class than are Anglo boys and girls. About 25 percent of the Anglo boys and 35 percent of the Anglo girls feel that the teachers do not practice what they preach, as compared to only 9.8 percent of the Latin boys and 9.2 percent of the Latin girls.

Latin boys have more school problems than Latin girls, and to a greater degree than Anglo boys, in getting low grades, in spelling and grammar, and in writing. The Latin boys also have more trouble in reading and making oral reports. The Latin girls worry more than Anglo girls about their grades.

The Anglo boys and girls do have some school problems to a greater de-

gree than do the Latins as Table 1 shows. About 25 percent of the Anglo boys feel that school is too strict as compared to only 15.4 percent of the Latin boys. Also, 23.8 percent of the Anglo girls feel there is too little freedom in classes as compared to only 10 percent of the Latin girls. A greater percentage of the Anglo girls are not interested in books and are bored with dull classes than are the Latin girls.

Tables 2 and 3 show the problems troubling students most. Eight out of the twelve problems troubling the Latin girls most are school problems. Only one of the seven problems troubling Anglo girls most is a school problem, and that is dull classes. The Latin girls' greatest problem is that of getting low grades in school. They also worry about arithmetic, fear to fail in school work, and feel that they are not smart enough. Four of the eight problems troubling the Latin boys the most are in the school area and indicate fears of failure and worry about grades. The Anglo boys have four problems in the school area, but not to the extent the Latins do.

School does present problems to both races, and it is up to the schools to help their students solve their problems. The Latin-Americans have more school-connected problems than do the Anglo-Americans, and since it is possible to identify many of the specific problems troubling them, it should be easier for the schools to help solve these problems.

The problems in the other six areas

TABLE 1. School Problems Showing at Least a 10 Percent Difference between Anglo- and Latin-American Boys and Girls

	Percent of Boys		Percent of Girls	
Problem	Anglo	Latin	Anglo	Latin
Getting low grades in school	35.2	62.8	23.8	43.3
Afraid of failing in school work	30.0	46.0	34.9	46.4
Trouble with arithmetic	33.8	59.0		
Trouble with spelling or grammar	20.0	32.0		
Not spending enough time in study	20.0	35.9		
Can't keep my mind on my studies	28.2	42.3		
Trouble with oral reports	20.0	32.2		
Not interested in certain subjects	38.0	28.2		
School is too strict	25.3	15.4		
Teachers not practicing what they preach	25.3	9.8	34.9	9.2
Slow in reading	15.5	29.5		
Trouble with writing	10.0	20.0		
Not smart enough	11.2	35.9	11.1	41.1
Afraid to speak up in class	15.5	32.0	20.6	49.5
Worried about grades			28.6	47.4
Not interested in books			28.6	6.2
Too little freedom in classes			23.8	0.0
Dull classes			20.6	11.4

TABLE 2. Problems of Most Concern Marked by at Least 10 Percent of Both Groups of Boys

Anglo Boys' Problem	Percent	Latin Boys' Problem	Percent
Worried about grades	21.1	Getting low grades in school	33.3
Too little chance to do what I want to do	19.7	Afraid of failing in school work	23.0
Getting low grades in school	14.1	Trouble with arithmetic	20.5
Not allowed to run around with kids I like	14.1	Lacking self-control	14.2
Wanting to earn some of my own money	14.1	Worried about grades	11.5
Don't get enough sleep	12.7	Needing a job during summer vacations	11.5
Bashful	12.7	Wanting to earn some of my own money	10.3
Trouble with spelling or grammar	12.7	Underweight	10.3
Restless to get out of school and into a job	11.3		
Needing to find a part-time job now	11.3		
Don't like school	10.0		
Not interested in certain subjects	10.0		

are not directly school problems, but should be of interest to counselors and others interested in helping students. The results will be summarized as briefly as possible.

In the Money, Work and Future area, Latin girls again have more problems and to a greater degree. Of the six problems in this area showing at least a 10 percent difference between the girls, the Latin girls have five to a greater degree. These are deciding what to take in high school, needing a job during vacations, having no regular allowance, having to ask parents for money, and not knowing how to look for a job. More of the Anglo girls had as a problem wondering if they would ever get married. The Anglo boys had one more problem in this area than did the Latin boys and that was wanting to know about trades.

The Anglo girls had more problems in the Self-centered Concerns area. More Anglo girls than Latin girls are bothered by being afraid of making mistakes, not being able to forget some mistakes they have made, and day-dreaming too much. The Latin girls are more concerned over forgetting things. A higher percentage of Latin boys than Anglos have the problems of lacking self-control, being lazy, and not taking some things seriously enough.

In the Boy and Girl Relations area, more Latin girls have the problems of learning how to dance and dating more frequently than do the Anglo girls. The boys do not show much difference here, but as problems of most con-

TABLE 3. Problems of Most Concern Marked by at Least 9.5 Percent of the Anglo Girls and 10 Percent of the Latin Girls

Anglo Girls' Problem	Percent	Latin Girls' Problem	Percent
Dull Classes	12.8	Getting low grades in school	23.7
Not allowed to have dates	11.1	Trouble with arithmetic	23.7
Parents not understanding me	9.5	Worried about grades	18.6
Wanting to earn some of my own money	9.5	Afraid of failing in school work	17.5
So often not allowed to go out at night	9.5	Afraid to speak up in class	13.4
No place to entertain my friends	9.5	Afraid of tests	12.4
Talking back to my parents	9.5	Often have headaches	11.3
		Wanting to earn some of my own money	11.3
		Learning how to dance	11.3
		Not smart enough	10.3
		Not good looking	10.3
		Trouble with oral reports	10.3

cern, the Anglo boys mark not being allowed to run around with kids they like, and too little chance to do the things they want to do. The Latin boys mark no problems to a significant degree in this area.

In the Home and Family area, the girls seem to have about the same frequency of problems. As problems troubling them most, however, one out of ten Anglo girls marked the problems of parents not understanding them and talking back to their parents, while the Latin girls marked no problems above the 10 percent level of frequency in this area. The Anglo boys gave as a problem that they wanted things their parents would not give them, but the Latin boys had no significant problems here.

In the Relations to People in General area, the Latin girls are more frequently concerned over being jealous, and the Anglo girls are more concerned with wanting a more pleasing personality. The Anglo boys are more concerned over being bashful and wanting a more pleasing personality. The Latin boys do not show any significant problems in this area.

The Health and Physical Development area has one problem in it bothering at least 10 percent more of the Latin girls than Anglo girls, and that is they often are not hungry for their meals. However, as a problem bothering them most in this area, 10.3 percent of the Latin girls often have headaches and feel that they are not good-looking. The Anglo girls marked no significant percentage of problems in this area. In this area many of the Latin boys are worried about being underweight, and this is also a problem bothering them most. As a matter of most concern to the Anglo boys they mark not getting enough sleep.

It is not the purpose of this study to

attempt to point out how the school can help its students solve their problems, as that can be done only by experts in the field. It is hoped that school counselors and administrators, by knowing what the main problems of the Anglo-American and Latin-American students are, can mobilize their counseling and curricular offerings to meet these problems. A program could be planned to meet these indicated problems through curricular and extracurricular means, taking into account some of the differences between Anglo- and Latin-American students. There are more likenesses than differences in the problems of the two groups, and these possibly should be dealt with first, but special attention could be given to the more frequent and acute school problems of the Latin Americans.

9.6 *Negro Academic Motivation and Scholastic Achievement*

ROBERT LEE GREEN AND WILLIAM W. FARQUHAR

Reprinted from the *Journal of Educational Psychology,* 1965, **56,** 241–243, with permission of the authors and the American Psychological Association, Inc. Robert Lee Green is the Associate Provost for Urban Affairs at Michigan State University. William W. Farquhar is a counseling psychologist on the faculty of the College of Education, Michigan State University.

Although verbal scales of intelligence tests are reputed to be the best single predictor of scholastic success for most children in the schools, there has long been a suspicion that they do not work well for some socially deprived students. This study by Green and Farquhar shows that verbal intelligence scores gave the usual correlations for Negro female high school students as well as for white students of both sexes, but were of no value for Negro males. The test that *did* predict school performance for Negro males, however, was a measure of academic motivation—a kind of academic need to achieve. Scores from the motivation test, incidentally, were also better predictors of academic success for female students of both races than were the more conventional verbal aptitude test scores. This study is of particular interest, not only because it shows that verbal intelligence tests are likely to be invalid as academic predictors for Negro males, but also because it points to the importance of motivational factors in school success.

Recent studies have indicated that the typical Negro student fails to achieve as well (Boykin, 1955; Bullock, 1950), drops out of school more frequently (Conant, 1961), and demonstrates a lower need for achievement than his Caucasian counterpart (Lott & Lott, 1963). Many educators assert that the school achievement (grade-point average) of both groups is related to achievement motivation and academic aptitude.

Norton (1959) found that the total Differential Aptitude Test (DAT) correlated significantly with science-achievement scores of white ninth-grade males and females. Jacobs (1959) found that the DAT Verbal Reasoning correlated significantly with grade-point average (GPA) for senior high school males and females.

Additional studies (Bennett, Seashore, & Wesman, 1959) have demonstrated that cognitive and personality factors correlate with achievement for Caucasian students. However, few have attempted to explore the relationship between the latter factors and achievement for Negro student populations.

The purpose of the present study was to investigate the relationship of personality and cognitive factors with academic achievement (GPA) for eleventh-grade Negro and white students of both sexes.

Method

Subjects

The Negro sample consisted of 104 males and 129 females selected from two Detroit-area high schools with a total eleventh-grade school enrollment of 700 students. The schools were selected a priori in order to represent a full range of socioeconomic environments.

The Caucasian sample, tested by the second author, consisted of 254 males and 261 females randomly selected from a population of 4,200 eleventh-grade students from nine high schools in eight Michigan cities.[1]

Measures

Three measures were gathered on both samples:

1. Michigan State M Scales—a theoretically based objective measure of academic motivation. The M scales consist of four subtests which were designed to assess the following motivational components: (*a*) the need for academic achievement (Generalized Situational Choice Inventory), (*b*) academic self-concept (Word Rating List), (*c*) occupational aspirations (Preferred Job Characteristics Scale), and (*d*) academic personality factors (Human Traits Inventory). The total scale contains 139 male and 136 female items.

2. Aptiture Measure—the verbal score of the School and College Ability Test (SCAT) and the Verbal Reasoning score of the DAT were obtained from the school records of the Negro and Caucasian students, respectively.[2]

[1] For a full description of the sample-selection procedure and the development of the Michigan State M scales, see Office of Education Cooperative Research Project No. 846.

[2] Both the SCAT and DAT Verbal tests correlate comparably with the American Council on Education (ACE) Linguistic test. SCAT Verbal and ACE Linguistic = .89, DAT Verbal and ACE Linguistic = .84 females, .74 males.

3. School Achievement (GPA)—each student's GPA was computed using ninth- and tenth-grade subjects. Only academic subjects were included, that is, those requiring homework.

Results

The correlations between achievement (GPA) and aptitude for both races and sexes are shown in Table 1. As indicated in this table, there is no correlation between verbal aptitude and achievement (GPA) for Negro males despite the significant correlation between verbal aptitude and GPA for Negro females. All motivation subtests —except the male Human Traits Inventory (HTI)—correlate significantly with achievement for both Negro males and females. The self-concept—Word Rating List (WRL)—is the best single prediction of achievement for the Negro sample.

Both verbal aptitude and the motivation scales correlate significantly with achievement for the white male and white female samples. The best single predictor of achievement is, for the white male sample, verbal aptitude (.62) and, for white females, the self-concept (WRL) scale (.34). The M-scale total correlates significantly with achievement for all groups.

Discussion

The most important finding of the study was the lack of correlation between aptitude and achievement (—.01) for Negro males. This finding is noteworthy in light of the correlation between aptitude and GPA (.64) for white males. The relationship between aptitude and achievement for Negro males must be qualified as pertaining to a northern urban educational system. It may be hypothesized that in a rural southern segregated educational system, verbal aptitude might again be a significant

TABLE 1. Correlations between Aptitude and M-scale Subtests with Grade-Point Average as a Function of Race and Sex

Sample		Aptitude and M-scale Subtests Correlated with Grade-Point Average					
Race and Sex	Verbal Aptitude	GSCI	HTI	PJCS	WRL	M total	*N*
Negro							
Male	—.01	.26*	.14	.30*	.36*	.37*	104
Female	.25*	.46*	.40*	.34*	.64*	.55*	129
White							
Male	.62*	.50*	.42*	.32*	.51*	.50*	254
Female	.21*	.21*	.29*	.18*	.34*	.43*	261

Note. Abbreviations used: GSCI, Generalized Situational Choice Inventory; PJCS, Preferred Job Characteristics Scale. The variability of verbal aptitude and the HTI subtest for the Negro Male sample did not exceed the variability of the other subtests.
*$p < .05$.

predictor for Negro males because, in most segregated educational systems, those scoring high on aptitude tests would be given priority on the typically meager educational facilities. Thus, in a segregated system verbal-aptitude results may become a self-fulfilling prophecy. However, in our northern sample, educational opportunity was available in sufficient abundance to reach a reasonably wide range of youngsters, irrespective of aptitude.

Among the subtests of the M scales, the self-concept scale (WRL) is the best predictor of achievement for Negro males (.36) and females (.64) and white females (.34). This finding supported by the recent research of Payne and Farquhar (1962) and indicates the strong relationship between the students' self-perception and school achievement.

For the Negro males, the M scales appear to be more valid predictors of achievement than verbal aptitude. This finding has implications for school administrators and counselors since scores on aptitude tests are often used in making student educational- vocational decisions. Because verbal aptitude was shown to be a poor predictor of achievement for the Negro male students of this sample, critical examination should be given to the validity of the SCAT Verbal and other verbal-aptitude measures before making decisions concerning Negro students solely on verbal measures.

The finding that verbal aptitude was a significant predictor of achievement for Negro females in contrast to Negro males should be more fully explored.

Obviously, cross validation of these correlations is needed. Furthermore, the value of the M scales in estimating achievement for this sample of Negro students emphasizes the relationship between nonintellectual factors and school performance. It may well be that many Negro students (especially males) are being graded on other than academic performance (e.g., social desirability). Future studies employing other forms of standard achievement tests are needed to isolate the pertinent factors which determine school achievement of this minority group.

Summary

Separate samples of 233 Negro and 515 Caucasian high school students of both sexes, randomly selected to represent a wide range of socioeconomic environments, were tested as to verbal aptitude, academic achievement, and academic motivation. Except for Negro males, both samples obtained significant correlations between verbal aptitude and achievement. The Negro males showed no such relationship between aptitude and achievement, but academic-motivation tests (the M scales) correlated significantly with achievement for all groups of interest.

References

Bennett, G. K., Seashore, H. G., and Wesman, A. G. *Differential Aptitude Test Manual.* (3rd el.) New York: Psychological Corporation, 1959.

Boykin, L. L. The reading performance of Negro college students. *Journal of Negro Education,* 1955, **24**, 435–441.

Bullock, H. A. A comparison of the academic achievements of white and Negro high

school graduates. *Journal of Educational Research,* 1950, **44**, 179–192.

Conant, J. B. *Slums and Suburbs.* New York: McGraw-Hill, 1961.

Jacobs, J. M. Aptitude and achievement measures in predicting high school academic success. *Personnel and Guidance Journal,* 1959, **37**, 334–341.

Lott, Bernice E., and Lott, A. J. *Negro and White Youth; a Psychological Study in a Border-State Community.* New York: Holt, Rinehart & Winston, 1963.

Norton, D. P. The relationship of study habits and other measures to achievement in ninth grade general science. *Journal of Experimental Education,* 1959, **27**, 211–217.

Payne, D. A., and Farquhar, W. W. The dimensions of an objective measure of academic self-concept. *Journal of Educational Psychology,* 1962, **53**,187–192.

9.7 Changes in Attitudes Toward Negroes of White Elementary School Students After Use of Multiethnic Readers

JOHN H. LITCHER AND DAVID W. JOHNSON

Reprinted from the *Journal of Educational psychology,* 1969, **60**, 148–152, with permission of the authors and the American Psychological Association, Inc. John H. Litcher is on the faculty of the College of Education of the University of Florida, and David W. Johnson is a member of the educational psychology faculty of the University of Minnesota.

The papers included in this section show how the educational problems of socially deprived youngsters are related to experiences during early childhood and at school. Still another factor is the prejudice of the larger society, which is particularly severe toward nonwhite and non-Anglo minorities. This prejudice has the effect of a "self-fulfilling prophecy," in the sense that minority group children, being the target for negative attitudes, come to expect little of themselves and accordingly perform more poorly than they would otherwise. The attitudes on which ethnic prejudice is based are learned, often taught unconsciously, and it makes sense to begin with children who will be the citizens of tomorrow. The study by Litcher and Johnson shows that introducing into the schools curricular material that recognizes the contributions of all ethnic groups can make some first steps toward reducing and eventually eliminating ethnic prejudice. Of special interest in this article is the description of measures used to assess the prejudice of second-graders.

Changing racial attitudes on a wide scale basis is one of the most important social-psychological problems of our society. It is evident that from a very early age white children are prejudiced against Negroes (Blake & Dennis, 1943; Goodman, 1948; Gregor & McPherson, 1966; Horowitz, 1936, 1939; Katz & Braly, 1933; Landreth & Johnson, 1953; Radke & Trager, 1950; Radke, Trager, & Hadassah, 1949). There is some empirical evidence which indicates that under certain conditions (which have not been adequately researched) the attitudes of whites toward Negroes may be changed through direct experience. Singer (1967), for example, in a recent study of the effects of integrated classrooms upon the racial attitudes of fifth-grade children, found that white children in integrated schools, compared with white children in segregated schools, are more accepting of Negroes and more familiar with Negro celebrities. The more intelligent the white child in the integrated school, the more favorable are his attitudes toward Negroes.

It is not possible, however, to provide every white child with direct experiences with Negroes. In Minnesota and North Dakota, for example, the Negro population is so small that such direct experiences are impossible. One alternative to direct experience with Negroes is exposure to materials which portray Negroes in a positive way, contradicting prevailing prejudices and stereotypes. Research in social perception (Allport & Postman, 1945; Bartlett, 1932) and in the learning of controversial material (Edwards, 1941; Jones & Aneshansel, 1956; Levine & Murphy, 1943) suggests that materials portraying Negroes positively would be either distorted in various ways to support the prevailing stereotypes and prejudices or ignored and quickly forgotten. Research in counter-conditioning (Bandura & Walters, 1963), however, would predict that such an approach would be effective. If, for example, the stimulus "Negro" (which elicits a negative response) is repeatedly paired with the cluster of stimuli characteristic of "middle class" (which elicits a positive response), the stimulus "Negro" will elicit the positive response associated with "middle class"—if the estimulus "Negro" does not elicit a more powerful response than the response elicited by the stimuli characteristic of "middle class".

An exploratory study was conducted contrasting the social perception, social learning, and the counter-conditioning hypotheses by investigating the effect of multiethnic readers upon the racial attitudes of white elementary students. Multiethnic readers are readers which contain characters from several different racial and ethnic groups. In the readers used, Negroes are portrayed as having middle class characteristics (works hard, dresses nicely, is clean, etc.) in integrated situations.

Method

This study employed a pretest-posttest control group design. Experimental groups used a multiethnic reader for 4 months while control groups used the traditional reader. Both groups were interviewed before and after the experimental treatment. The study was conducted from February 1967 through May

of the same year. Sixty-eight white, middle-class children were studied, 34 classified by their teachers as upper group readers and 34 classified as middle group readers. Both the multiethnic and the regular second-grade readers were used by each teacher in each classroom. Eight reading groups in four second-grade classrooms in two public elementary schools participated in the study. Through random assignment, two classrooms (one in each school) used the multiethnic reader in their upper reading group and the regular reader in their middle reading group. The other two classrooms (one in each school) used the regular reader in their upper reading group and the multiethnic reader in their middle reading group.

The four teachers who participated in the study were randomly selected from volunteers within the school system. The teachers' interest in this study was prompted by the opportunity it offered them to participate in research. They were generally informed as to the nature of the study and asked to teach the experimental and control groups as similarly as possible. Since the basal approach to reading instruction was followed, the teaching included the development of work recognition skills, comprehension skills, reading skills in other content areas, oral and silent reading, and emphasis on a personal reading program. A record of any discussion relating to race relations was requested. According to their reports, the teachers did not at any time initiate a discussion of the fact that many of the characters in the multiethnic reader were nonwhite. Neither did they encourage student discussion of the racial differences of the characters in the reader while the study was in progress. The students commented very little on the differences in race of the characters about whom they were reading. The multiethnic readers were the only multiethnic materials in the classroom.

The Scott-Foresman multiethnic (Robinson, Monroe, & Artley, 1965) and regular (Robinson, Monroe, & Artley, 1963) second-grade readers were used in the study. These readers are identical except for the pictures (some of the characters in the pictures in the multiethnic reader are nonwhite) and the names used to represent the characters of the racial and ethnic groups found in the readers.

For both the pretest and the posttest each child was interviewed individually. Four tests were presented in random order and all were given in one sitting. On the average it took 9 minutes to administer the tests. All questions were asked of each child, but answers were not made compulsory.

During the 4 months of the experimental treatment three children, all in the experimental group, moved out of the school district. They were, therefore, not available for the posttest.

The study was conducted in a Midwestern city of 50,000 inhabitants. The total Negro population in this city is less than 100. Of the 6,181 children attending the city's elementary schools, 10 are Negroes. No Negro children attended the two elementary schools studied.

The instruments used in this study were a variation of the Clark Doll Test (Gregor & McPherson, 1966), the Horowitz and Horowitz (1938) "Show Me" and Categories tests, and a Direct

Comparison test (Blake & Dennis, 1943). In the Clark Doll Test the children were presented with two dolls which differed only in skin color (one white, one dark brown). The children were asked to point to one doll as a response to the following questions: Show me the doll that (*a*) you would like to play with, (*b*) you like best, (*c*) is a nice doll, (*d*) has a nice color, (*e*) looks bad, (*f*) looks like a white child, (*g*) looks like a colored child, (*h*) looks like a Negro child, and (*i*) looks like you.

The "Show Me" test developed by Horowitz and Horowitz (1938) consists of 12 portraits (3 white boys, 3 white girls, 3 Negro boys, and 3 Negro girls) placed randomly on a large sheet of paper, The following questions were asked of the children: Please show me the one that (*a*) you'd like to sit next to at school, (*b*) you'd want to play with, (*c*) comes from a poor home, (*d*) you do not want in your school, (*e*) you'd like to have as your cousin, (*f*) doesn't look very smart, (*g*) you would want to come to your house for a long visit, and (*h*) you do not like.

In the Categories test (Horowitz & Horowitz, 1938), five pictures mounted on a page were presented to each child. The children were asked to reject one picture as not belonging to the group. Categories of race versus sex and race versus age were used. For example, one page might contain five pictures, three white boys, one white girl, and one Negro boy. If the Negro boy was selected as not belonging, race is a more salient category than sex for that child, If the white girl was selected, sex could be considered as the more salient category. The test was designed to analyze the strength of race, sex, and age categories for the children.

The Direct Comparison test (Blake & Dennis, 1943) required the children to make direct comparisons between Negroes and whites in regard to 18 traits. The children were asked to indicate whether the trait was more characteristic of whites or Negroes or to respond "no difference" or "don't know." Examples of the traits used are: cheerful, honest, lazy, forgetful, neat, clean, lies.

Results

As part of the pretest each child was asked to respond to several questions dealing with racial identification. From Table 1 it may be seen that with the exception of one child, all the children correctly identified the dolls used in the Clark Doll Test with their appropriate racial group. From these data it may be concluded that with one possible exception, all the children studied were able to respond to questions dealing with racial membership.

Each child studied responded to four tests of racial attitudes. Since the investigators were interested only in general changes in racial attitudes resulting from the use of multiethnic readers and not in responses to the specific questions of each test a general score of favorableness of attitudes toward Negroes was derived for each child for each test. A child was given one point for each response which indicated favorable attitudes toward Negroes. For the Clark Doll Test the range of possible scores was 0–5, for the "Show Me"

TABLE 1. Frequency of Responses of White Second-Grade Children to Items of Race Identification on the Clark Doll Test–Pretest

Which Doll	Multiethnic Reader Group [a]		Regular Reader Group [b]	
	Negro Doll	White Doll	Negro Doll	White Doll
Looks like a white child	1	30	—	34
Looks like a colored child	31	—	34	—
Looks like a Negro child	31	—	34	—
Looks like you	1	30	—	34

[a] $N = 31$.
[b] $N = 34$.

test the range of possible scores was 0–8, and for the Categories test, 0–6. For the Direct Comparison test the proportion of responses favorable to Negroes was used, the possible range of scores being from 0–100%.

In order to see if the tests were independent of each other the responses of the children to the four tests were correlated. From Table 2 it may be seen that the tests were only slightly correlated and, therefore, the data for each were analyzed separately.[1]

On all tests there were no significant differences between the experimental and control group on the pretests. In

[1] Table 2 has been omitted in the interest of brevity.

TABLE 3. Comparison of Attitudes toward Negroes of White Second-Grade Children–Posttest

Test	Multi-ethnic Reader Group [a]	Regular Reader Group [b]	Significance
Clark Doll Test (range: 0–5)	1.39	0.44	$F = 15.90$; $p < .0002$
Show Me Test (range: 0–8)	2.58	1.47	$F = 8.71$; $p < .005$
Categories Test (range: 0–6)	4.42	3.44	$F = 4.38$; $p < .04$
Direct Comparison Test	51%	24%	$F = 14.94$; $p < .0003$

Note. The higher the score the more favorable the attitudes toward Negroes. An analysis of covariance was used in alalyzing the data.
[a] $N = 31$.
[b] $N = 34$.

order to control for slight differences between the groups, however, an analysis of covariance was used in analyzing the results of the posttest. The data in Table 3 indicate that on all four tests the children using the multiethnic readers responded significantly more favorably toward Negroes than the children using the regular readers.

Discussion

The results of this study dramatically indicate that the use of multiethnic readers in an elementary school will result in more favorable attitudes toward Negroes. The data from the Clark Doll Test indicate that the use of the multiethnic reader decreased the preference for one's own racial group over the other. While the control group expressed marked preferences for the white rather than the Negro doll, the experimental group *S*s were far less unanimous about their preferences.

In a recent study on the Clark Doll Test, Greenwald and Oppenheim (1967) report that 19% of the white children they interviewed (taking their more conservative figure) identified the Negro doll in response to the question, "Show me the doll that looks like you." In an earlier study Morland (1963) found that 14% of the white children interviewed responded similarly. Greenwald and Oppenheim (1967) concluded on the basis of these findings that the amount of Negro misidentification found in the Clark and Clark (1940) and other similar studies is misleading and they did not have a control group of white children. The present study (1.5% white misidentification) and the study of Gregor and McPherson (1966; 0% white misidentification) give no support to their findings or their conclusion.

On the "Show Me" test, the use of the multiethnic reader resulted in a reduction of the amount of social distance placed between the white and Negro racial groups. On the Categories test, the children in the experimental group were less likely to exclude a child on the basis of race than were the controls. The data for the Direct Comparisons test, furthermore, indicate that the experimental *S*s were less likely to attribute negative traits to Negroes and positive traits to whites than were the control *S*s. Examination of the individual items revealed that the experimental group basically became equalitarian in their response. Johnson (1967) found that Negro children who were taught Negro history in a Freedom School became much more convinced that Negroes and whites are equal. Thus, the use of the multiethnic reader had much the same effect on white children as learning Negro history had upon Negro children.

The evidence is quite clear. Through the use of a multiethnic reader, white children developed markedly more favorable attitudes toward Negroes. Under the conditions of this experiment, this finding supports the counter-conditioning hypothesis and does not support the social-perception and social-learning hypothesis. The implications of this finding hardly need elaboration. While it is not possible, due to lack of material resources and the distribution of the Negro population in the United States, for every white child to have direct experiences with Negroes (al-

though the investigators believe it is desirable), it is possible to increase the visibility of the Negro in the curriculum materials of the schools. Such an action should, through the reduction of prejudice, increase racial harmony.

A limitation on the generalization of the results of this study should be noted. The Negro population in the city in which this study was conducted is quite small (Negroes make up less than .2% of the total population of the city). The probability is very high that the children participating in the study had no direct experience with Negroes and that the Negro community does not represent an economic or social threat to the white community. The racial attitudes of the children studied, therefore, are probably not firmly rooted in direct experiences or reference group norms.

Summary

This study investigated the effect of curriculum materials which portray Negroes in a way which is contradictory to prevailing prejudices and stereotypes upon the attitudes toward Negroes of white second-grade school children in a Midwestern city. A pretest-posttest design controlling for the teacher, the classroom, the school, and the reading ability of *S*s was used. The 34 children in the experimental groups used a multiethnic reader which included characters from several different racial and ethnic groups for 4 months, while the 34 children in the control groups used the regular reader which included only whites. Use of the multiethnic reader resulted in marked positive change in *S*s' attitudes toward Negroes, supporting the counter-conditioning hypothesis.

References

Allport, G. W., and Postman, L. J. The basic psychology of rumor. *Transactions of the New York Academy of Sciences,* Series II, 1945, **8**, 61–81.

Bandura, A., and Walters, H. *Social Learning and Personality Development.* New York: Holt, Rinehart & Winston, 1963.

Bartlett, F. C. *Remembering.* Cambridge, England: Cambridge University Press, 1932.

Blake, R., and Dennis, W. The development of stereotypes concerning the Negro. *Journal of Abnormal and Social Psychology,* 1943, **38**, 525–531.

Clark, H. B., and Clark, M. K. Skin color as a factor in racial identification of Negro preschool children. *Journal of Social Psychology,* 1940, **11**, 160.

Edwards, A. L. Political frames of reference as a factor influencing recognition. *Journal of Psychology,* 1941, **36**, 34–50.

Goodman, M. E. Evidence concerning the genesis of interracial attitudes. *American Anthropologist*, 1946, **48**, 624–630.

Greenwald, H. J., and Oppenheim, D. B. Reported magnitude of self-misidentification among Negro children—artifact? *Journal of Personality and Social Psychology,* 1968, **8**, 49–52.

Gregor, A. J., and McPherson, D. A. Racial attitudes among white and Negro children area. *The Journal of Social Psychology,* 1966, **68**, 95–106.

Horowitz, E. The development of attitude toward the Negro. *Archives of Psychology,* 1936, **194**, 5–47.

Horowitz, R. E. Racial aspects of self-identification in nursery school children. *Journal of Psychology,* 1939, **7**, 91–99.

Horowitz, L., and Horowitz, R. E. Development of social attitudes in children. *Sociometry,* 1938, **1**, 301–339.

Johnson, D. W. The effects of a freedom school on its students. In R. Dentler, B. Mackler, and M. E. Warshauer (Eds.), *The Urban R's: Race Relations as the Problem in Urban Education.* New York: Praeger, 1967.

Jones, E. E., and Aneshansel, J. The learning and utilization of contravaluant material. *Journal of Abnormal and Social Psychology,* 1956, **53**, 27–33.

Katz, D., and Braly, K. W. Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology,* 1939, **28**, 280–290.

Landreth, C., and Johnson, B. C. Young childrens' responses to a picture and insert test designed to reveal reactions to persons of different skin color. *Child Development,* 1953, **24**, 63–80.

Levine, J. M., and Murphy, G. The learning and forgetting of controversial material. *Journal of Abnormal and Social Psychology,* 1943, **38**, 507–517.

Morland, J. K. Racial self-identification: a study of nursery school children. *American Catholic Sociological Review,* 1963, **24**, 231–242.

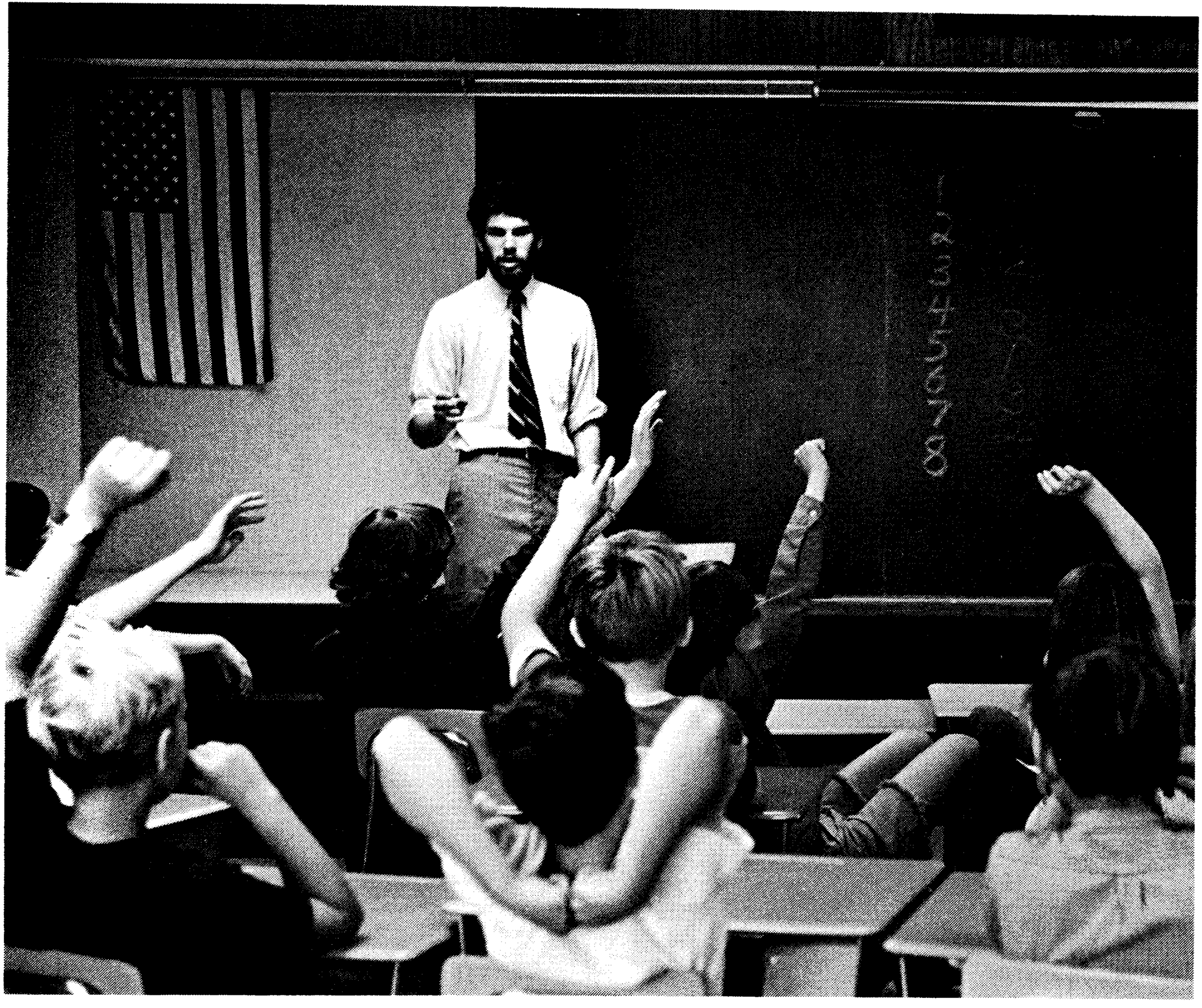
Radke, M. J., and Trager, H. G. Children's perception of the social roles of Negroes. *Journal of Psychology,* 1950, **29**, 3–33.

Radke, M. J., Trager, H. G., and Hadassah, D. Social perceptions and attitudes of children. *Genetic Psychology Monograph,* 1949, **40**, 327–447.

Robinson, H. M., Monroe, M., and Artley, A. S. *New More Friends and Neighbors 2-2* and *New Friends and Neighbors 2-1,* Chicago: Scott, Foresman, 1963.

Robinson, H. M., Monroe, M., and Artley, A. S. *More Friends Old and New 2-2* and *Friends Old and New 2-1,* Chicago: Scott, Foresman, 1965.

Singer, D. The influence of intelligence and an interracial classroom on social attitudes. In R. Dentler, B. Mackler, and M. E. Warshauer (Eds.), *The Urban R's: Race Relations as the Problem in Urban Education.* New York: Praeger, 1967.



Sybil Shackman from Monkmeyer

PART 10

The Psychology of Being a Teacher

10.1 Teachers' Communication of Differential Expectations for Children's Classroom Performances: Some Behavioral Data

JERE E. BROPHY AND THOMAS L. GOOD

Reprinted from the *Journal of Educational Psychology,* 1970, **61**, 365–374, with permission of the authors and the American Psychological Association, Inc. Jere E. Brophy and Thomas L. Good are both on the educational psychology faculty of the University of Texas.

Several of the articles in the first nine parts of this book have taken note of a phenomenon termed "the self-fulfilling prophecy," a term used to describe situations in which individuals who have certain expectations act in ways that are likely to confirm them. They usually are aware of their expectations, but they are generally unaware of the way in which they make certain that their expectations will be achieved, and, as a consequence, attribute the outcome to others, to fate, or whatever. An example of the self-fulfilling prophecy is the poor school performance of socially disadvantaged children. Teachers expect poor performance of them and communicate this expectation in many ways. The children similarly make "self-fulfilling prophecies" and act to confirm their expectations that they cannot learn. In the present article, Brophy and Good examine the effect of teacher expectations on their behavior and show how they reinforce and encourage the striving of pupils for whom they have high expectations and ignore the efforts of those for whom they have low expectations.

Rosenthal and Jacobson (1968) assert on the basis of controversial research presented in *Pygmalion in the Classroom* that teachers' expectations for student performance function as self-fulfilling prophecies. The "expectancy effects" in the Oak School experiment described in *Pygmalion* are not as inconsistent as the authors' interpretations of them would suggest, however, and even the support that they do provide is questionable on methodological grounds (Barber & Silver, 1968; Snow, 1969; Thorndike, 1968). Even if the data and their interpretation are accepted, the Rosenthal and Jacobson work remains only a demonstration of the *existence* of expectancy effects; their study did not address itself to any of the events intervening between the inducement of teacher expectations and the administration of the criterion achievement test. The present study focuses on these intervening processes, applying the method of classroom interaction analysis to identify and document differential teacher expectations to individual children.

The lack of data concerning the causal mechanisms at work in the Rosenthal and Jacobson study, combined with the tendency in most secondary sources to oversimplify or exaggerate their findings has cast an aura of magic or mystery around expectation effects. Consequently, it is important to conceptualize such phenomena as outcomes of observable sequences of behavior. The explicit model assumed in the present research may be described as follows: (*a*) The teacher forms differential expectations for student performance; (*b*) He then begins to treat children differently in accordance with his differential expectations; (*c*) The children respond differentially to the teacher because they are being treated differently by him; (*d*) In responding to the teacher, each child tends to exhibit behavior which complements and reinforces the teacher's particular expectations for him; (*e*) As a result, the general academic performance of some children will be enhanced while that of others will be depressed, with changes being in the direction of teacher expectations; (*f*) These effects will show up in the achievement tests given at the end of the year, providing support for the "self-fulfilling prophecy" notion.

A series of interrelated studies will be required to investigate systematically the full model from beginning (how do teachers form differential expectations in the first place?) to end (how do children change so as to begin to conform more closely to teacher expectations?). The present study deals with the second step: given differential teacher expectations, how are they communicated to the children in ways that would tend to cause the children to produce reciprocal behavior? To begin to answer this question, the present study approached the problem through classroom interaction analysis. In contrast to the usual classroom interaction study, however, the present research focused on dyadic interaction between the teacher and individual children.[3]

[3] In the study of dyadic interaction of the individual child (or teacher-child dyad) becomes the unit of analysis, rather than the class as a group. For a discussion of the advantages of this method for studying traditional teacher-effectiveness variables and of applications of the method to problems that cannot be approached through ordinary interaction analysis methods, see Good and Brophy (1971).

Method

Subjects

The research was carried out in four first-grade classrooms in a small Texas school district which serves a generally rural and lower-class population. However, a large military base located within the district contributes about 45% of the students in the school in which observations were taken. Children from the base tend to be from more urban backgrounds and of a somewhat higher socio-economic status than the local children. The ethnic composition of the school is about 75% Anglo-American, 15% Mexican-American, and 10% Afro-American, which is representative of the general population of the area.

Research was carried out in four of the nine first-grade classrooms in the school, chosen because there were no assistant teachers present to complicate the picture (the other five classrooms had preservice teacher interns assisting the head teacher). The four teachers involved were asked to rank the children in their class in the order of their achievement. These instructions were deliberately kept vague to encourage the teachers to use complex, subjective criteria in making their judgments. The rankings were then used as the measure of the teachers' expectations for classroom performance for the children in their classes. In each class, three boys and three girls high on the teacher's list (highs) and three boys and three girls low on the teacher's list (lows) were selected for observational study. The highs were simply the first six eligible children on the list. This was generally true also for the lows, although a few children low on the lists were excluded from the study because they could not speak English fluently or because of suspected emotional or biological disturbance. Substitutes for each type of child (high boys, high girls, low boys, low girls) were also identified and these were individually observed on days when children in the designated sample were absent.

The teachers had been told that the study was concerned with the classroom behavior of children of various levels of achievement. They were not informed that their own behavior as well as that of the children was being specifically observed. Furthermore, the teachers thought that observations were being taken on everyone in the class and did not know that specific subgroups had been selected for study. By selecting subjects from the extremes of the distributions of teachers' rankings, the chances of discovering differential teacher treatment of the students were maximized. However, the school practiced tracking, achieving homogeneity within the nine classrooms by grouping the children according to readiness and achievement scores. Thus, at least in terms of test scores, objective differences among the children (and, therefore, objective support for the validity of teacher expectations) was minimized.

Observation system

Since the object of the research was to focus on differential treatment of different children, the observation system developed was addressed only to dyadic contacts between the teacher and an individual child, with lecture-dem-

onstration and other teacher behavior directed to the class as a group being ignored. Although the types of interactions coded were partly dictated by the range of situations seen in pilot studies, certain features of the coding system were built in for their specific relevance to the study of communication of differential teacher expectations. One major and consistent feature was that the source of the interaction was always coded, so that it could be determined later whether the interaction was initiated by the teacher or by the child. The types of dyadic interactions coded included teacher-afforded response opportunities and other types of interactions initiated by the children. The teacher-afforded response opportunities included recitations and reading turns in the reading groups and answers to teacher questions (coded separately as to whether they were *open* questions directed to the class as a whole or *direct* questions aimed at a particular child). Response opportunities were important events for studying teacher expectations, since at these times the children were attempting to deal with problems relevant to academic subject matter. Consequently, the sequential nature of the initiation and reaction cycles involved in them was retained in the coding system. In addition to coding response opportunities separately by type (as listed above), coders also noted the quality of the child's response (correct, incomplete or partially correct, incorrect or no response) and the type of feedback given by the teacher (praise, criticism, supplying the answer, repeating the question, rephrasing the question or giving a clue, or giving no feedback at all). This retention of sequential relationships allowed later analysis of the relative as well as the absolute differences between the groups.

All the dyadic contacts other than response opportunities as defined above were coded as either teacher-afforded communications (individual feedback regarding seat work or homework, requests that the child perform procedural or caretaking functions, and disciplinary action or evaluative comment about the child's behavior) or interactions initiated by the child (calling out answers, showing work to the teacher or asking questions about it, and seeking permission or other procedural contact). Sequential data were also built into the coding of these interactions. In addition to coding the type (academic, procedural, or disciplinary) and initiator (teacher or child) of the interaction, coders also kept track of the evaluative nature of the teacher's feedback (praise, criticism, or impersonal feedback). The terms "praise" and "criticism" referred to teacher reactions which went beyond the level of simple affirmation or negation or corrective feedback by complimenting or criticizing the child personally. Simple affirmation ("yes," "OK," "that's right") was not considered "praise" unless accompanied by obvious expression or gesture connoting excitement or warmth. The latter reactions were considered "praise," as were the words "good" and "fine," as well as other, more obvious forms of verbal praise. Similarly, simple negation ("no," "that's not it") was not considered "criticism," unless accompanied by expressions or ges-

tures communicating anger or disgust. In addition to the latter responses, verbal statements such as "that's a stupid answer," or "what's the matter with you?" were coded as "criticism." Most teacher feedback involved simple affirmation or negation and/or communication of information and was coded as "impersonal feedback" to distinguish it from praise and criticism. The fourth category, "no feedback," was coded if the teacher did not react in any way to the child's response and simply moved on to something else.

In addition to the coding of dyadic interactions as described above, the hand-raising behavior of the children was tallied as a measure of their tendency to seek response opportunities. This was coded after open questions, when the children raised their hands seeking to be called on to answer the question, and after some direct questions, when children raised their hands if the child called upon to answer the question gave a wrong answer or was unable to respond.

After several pilot applications in which the system was perfected and satisfactory intercoder reliability was established, observations were made on 4 separate days in each of the four classes. To equalize the time spent in each classroom and insure that the full range of classroom activities was included, the observation period extended for an entire morning or an entire afternoon (two of each for each class). Data were recorded for all periods of academic activity during the observation period, using one data sheet for the reading group and another for all other situations. No data were recorded when the class was out of the room for recess or washroom breaks. During nonacademic procedural activities (clean-up, getting in line, pledge to the flag, etc.), only disciplinary actions and behavioral evaluations were coded.

Data were recorded by two observers seated at the rear of the classroom. The observers were thus in front of the teacher but behind or to the side of the majority of the children, who were seated at small tables of six or eight. During each observation, one observer coded the interactions involving the six highs and the other coded the six lows. It had originally been intended that assignment of children to observers would be determined by seating location, since coding could be done more conveniently when the target children are seated close together. However, in three of the four classrooms the children were seated in order of achievement level (a fact which is itself a correlate of expectancy effects, as will be pointed out below), so that observations were made on intact high or low groups. Each observer's assignments were balanced between the high and low groups to eliminate the possibility that any obtained differences between expectancy groups could be attributed to observer differences.

Data analysis

A variety of measures were derived from the raw coding sheets through simple arithmetic procedures, and scores were assigned to each of the 48 individual subjects. Analyses of variance then were performed to assess the effects of teacher expectancy, sex,

and classroom (teacher) and their respective interactions on the obtained scores. Two types of measures were identified. The first, subsuming most of the simple frequency counts, involved group differences which are attributable to objective differences in the groups of children themselves. Consequently, any significant group differences discovered in these variables, while important in themselves, could not be taken as evidence of expectancy effects. The second set of measures, mostly percentage figures in which absolute frequency differences are statistically controlled in order to allow a comparison of relative differences between the groups, are interpreted as measures of expectancy effects. Teacher behavior tapped by these measures is more proactive or teacher initiated, going beyond simple reaction by the teacher to stimulation by the child. The distinction between these two types of measures is exemplified in the Results section below, in which the two types of findings are separately presented.

Results

The results are presented in three tables, each giving mean values for the four classes, two sexes, and two expectancy groups and the *p* values for Group Effects in Class × Sex × Expectancy analyses of variance. Although no predictions were made concerning differences by class, the data are presented to show the degree to which the teachers varied on the measures taken. In addition, any interactions of class variation with expectancy effects would affect the interpretation of the latter, and need to be investigated whenever they occur. Inspection of all three tables reveals that a significant class effect was obtained for the great majority of the variables. The greatest class variation occurs on the simple frequency columns, especially in Table 1, although class effects still usually reach significance even in the ratio measures related to teacher expectations presented in Table 3. Because of this large variation across classes and the frequent significant interaction of class with expectancy, the nature of the interaction was specifically investigated for each variable to determine the consistency of expectancy effects. This information is integrated into the discussion of expectancy effects below.

Tables 1 and 2 contain the data from variables measuring objective differences among the children or aspects of teacher-child interaction which cannot be unambiguously interpreted as due either to teacher expectation effects or to objective differences among the children. Data from variables which do appear to be independent of differences among the children and, therefore, interpretable as indexes of expectation effects are presented in Table 3.

Measures of the quantity and type of teacher-child contacts are shown in Table 1. Other than the large class differences, the data are most notable for the consistency of expectancy group differences on variables measuring the tendency to seek out the teacher and initiate contact with her. Children for whom the teacher held high expectations (highs) raised their hands

TABLE 1. Group Differences in Quantity and Type of Contacts with Teachers

Variable	Group *Ms*								*p* Values for Group Effects						
	Class				Sex		Expectancy		Class	Sex	Expectancy	Class × Sex	Class × Expectancy	Sex × Expectancy	Class × Sex × Expectancy
	1	2	3	4	Boys	Girls	Lows	Highs							
Number of times child raises hand	.00 [a]	4.00	25.75	21.33	12.00	13.54	8.88	16.67	.001	*ns*	.01	*ns*	.01	.05	.10
Number of times child is called on to answer a question	.00 [a]	.67	4.08	2.58	2.13	1.54	1.71	1.96	.001	*ns*	*ns*	*ns*	*ns*	.05	*ns*
Procedural interaction initiated by child	5.00	1.58	7.08	2.92	4.92	3.38	3.17	5.13	.001	.05	.05	.05	*ns*	*ns*	.01
Work-related interactions initiated by child	4.17	1.67	9.58	2.92	5.29	3.88	1.79	7.38	.05	*ns*	.01	*ns*	.05	*ns*	*ns*
Teacher-initiated procedural interactions	4.58	2.75	1.25	.67	2.50	2.13	2.58	2.04	.001	*ns*	*ns*	*ns*	.05	*ns*	*ns*
Teacher-initiated work-related interactions	9.83	4.50	4.58	.67	6.42	3.38	6.00	3.79	.001	.05	*ns*	.05	.05	*ns*	*ns*
Teacher-afforded behavioral criticisms	3.25	1.58	5.92	3.17	5.25	1.71	4.92	2.04	.05	.01	.01	*ns*	*ns*	.01	*ns*
Teacher-afforded questions during reading groups	4.50	2.58	3.75	5.83	4.63	3.46	4.79	3.29	*ns*	*ns*	.10	*ns*	.01	*ns*	*ns*
Calling out answers during reading groups	8.08	1.53	2.83	.50	3.58	2.92	2.96	3.54	.001	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Total teacher-afforded response opportunities	10.33	6.00	12.17	14.00	11.75	9.50	10.96	10.29	.001	.05	*ns*	*ns*	*ns*	*ns*	*ns*
Total child-initiated response opportunities	17.25	4.83	19.50	6.33	13.79	10.17	7.92	16.04	.001	*ns*	.01	*ns*	.05	*ns*	*ns*
Total dyadic contacts	46.17	20.08	45.75	25.67	41.33	27.50	33.67	35.17	.001	.01	*ns*	*ns*	*ns*	*ns*	*ns*

[a] There were no open questions in Class 1.

TABLE 2. Group Differences in Academic Performance and Teacher Evaluation

Variable	Group *Ms*								*p* Values for Group Effects						
	Class				Sex		Expectancy								
	1	2	3	4	Boys	Girls	Lows	Highs	Class	Sex	Expectancy	Class × Sex	Class × Expectancy	Sex × Expectancy	Class × Sex × Expectancy
Total correct answers	8.67	4.75	9.33	8.42	9.13	6.46	6.67	8.92	.05	.05	.10	*ns*	*ns*	*ns*	*ns*
Total part-correct, incorrect, or "don't know" responses	4.33	2.08	3.58	4.00	3.88	3.13	4.63	2.38	*ns*	*ns*	.01	*ns*	*ns*	*ns*	*ns*
Total problems in reading during reading group	16.75	1.46	11.75	6.83	9.38	9.02	13.04	5.35	.001	*ns*	.001	*ns*	.05	*ns*	*ns*
Average number of reading problems per turn during reading group	3.00	1.38	6.03	3.39	3.83	3.00	4.67	2.23	.01	*ns*	.01	*ns*	*ns*	*ns*	*ns*
Total number of times praised by teacher/total dyadic contacts	9.17	7.33	10.00	3.25	7.71	7.17	3.88	11.00	.05	*ns*	.001	*ns*	.10	*ns*	*ns*
Total number of times criticized by teacher/total dyadic contacts	16.58	13.42	23.75	16.42	22.88	12.21	24.33	10.75	.10	.001	.001	*ns*	.05	.05	*ns*
Criticism/Praise + Criticism	61.92	68.83	63.33	82.38	73.00	65.23	85.04	53.19	*ns*	*ns*	.001	*ns*	*ns*	*ns*	*ns*
Average SAT score (grade level equivalent)	1.79	1.57	13.9	1.73	1.58	1.66	1.42	1.82	.001	*ns*	.001	*ns*	*ns*	*ns*	*ns*

TABLE 3. Group Differences on Variables Related to the Communication of Teacher Expectations

Variable	Group *Ms*								*p* Values for Group Effects						
	Class				Sex		Expectancy		Class	Sex	Expectancy	Class × Sex	Class × Expectancy	Sex × Expectancy	Class × Sex × Expectancy
	1	2	3	4	Boys	Girls	Lows	Highs							
Number of direct questions from teacher	.42	2.00	2.25	4.00	2.67	1.67	1.83	2.50	.01	.10	*ns*	*ns*	*ns*	*ns*	*ns*
Number of times called on to answer open questions/ number of times child raises hand	[a]	.17	.16	.12	.18	.14	.20	.12	*ns*	*ns*	.05	*ns*	*ns*	*ns*	*ns*
Percentage of correct answers followed by teacher praise	6.25	7.17	19.25	3.25	12.83	5.13	5.88	12.08							
Percentage of wrong answers followed by teacher criticism	34.08	12.46	2.08	1.83	15.50	9.73	18.77	6.46	.001	*ns*	.01	.001	*.05*	*ns*	*ns*
Percentage of wrong answers followed by repetition or rephrasing of the question	9.75	18.71	14.79	33.88	23.88	14.69	11.52	27.04	*ns*	*ns*	*.10*	*ns*	*ns*	*ns*	*ns*
Percentage of reading problems followed by repetition or rephrasing of the question or by giving a clue	75.08	45.75	38.91	44.64	52.10	53.39	38.37	67.05	.05	*ns*	.01	*ns*	.01	*ns*	*ns*
Percentage of answers (correct or incorrect) not followed by any feedback from the teacher	2.50	8.92	10.58	14.17	8.00	10.08	14.75	3.33	.10	*ns*	*.001*	*ns*	*.05*	*ns*	*ns*

[a] There were no open questions in Class 1.

more frequently and initiated more procedural and especially more work-related interactions than did children for whom the teachers held low expectations (lows). The Class × Expectancy Group interactions with regard to child-initiated contacts reflect degree rather than direction of effect. The highs exceeded the lows in each class for hand raising, initiating work-related interactions, and total child-initiated response opportunities (the hand-raising effect excludes Class 1, where it could not be assessed because the teacher never asked open questions while her class was being observed). The highs also exceeded the lows in three of the four classes in initiating procedural interactions. There was a negligible reversal in Class 2, where this type of interaction was very infrequent (highs averaged 1.50, lows averaged 1.67). The only exception to the pattern of significant differences between highs and lows in child-initiated interactions occurred in the measure of calling out answers in the reading groups. The mean difference is in favor of the highs, but it is not a significant difference and the effect occurred in only one of the four classes. The data for child-initiated contacts may be summarized, then, in the statement that, outside the reading group at least, the highs seek out the teacher and initiate interactions with her more frequently than the lows. The difference is especially notable in work-related interactions: the highs much more frequently show their work to the teacher or ask her questions about it, and they initiate many more response opportunities.

The data for contacts initiated or controlled by the teacher are less clear than for those initiated by the children. The highs were called on more frequently to answer open questions, but the teachers initiated more procedural and work-related interactions with the lows and afforded them slightly more response opportunities. None of these differences reach significance, however. The only significant difference occurred with teacher-afforded behavioral criticisms, which more frequently went to the lows than the highs. This effect showed an important interaction with sex, due to the high frequency of teacher criticisms directed at boys in the low group. Males in the low group averaged 8.25 teacher behavior criticisms, as compared with 2.25 for boys in the high group (the corresponding figures for girls are 1.58 and 1.83). Sex also interacted with expectancy in the measure of hand raising, and again the boys in the low group were notably different from the other three groups. These boys averaged 6.25 on the hand-raising measure as compared to 17.75 for the boys of the high group (corresponding figures for the girls are 11.50 and 15.58).

The data regarding interactions initiated or controlled by the teacher may be summarized as follows: there is a tendency for the teachers to initiate more contacts with the lows than with the highs, but the teachers cannot be said to have been compensating for the superiority of the highs in child-initiated contacts because the trend is not completely consistent and because the only significant differences occur with teacher criti-

cisms rather than with work-related contacts or provision of response opportunities. While the data for child-initiated contacts showed strong expectancy group differences, the measures of teacher-initiated interactions were much more closely related to sex than to expectancy. Boys were higher than girls on all measures of teacher-initiated contacts, significantly so for work-related interactions, behavioral criticisms, and total teacher-afforded response opportunities. When teacher-child dyadic contacts of all types are totaled, a clear difference favoring boys is evident; there is no difference between expectancy groups. Differences between the highs and the lows are in quality rather than quantity of interaction with the teacher.

Group differences in quality of academic performance and in frequencies of teacher praise and criticism are presented in Table 2. Consistent expectancy group differences appear for all the variables in this table. The highs produced more correct answers and fewer incorrect answers than the lows, had fewer problems in the reading groups, and achieved higher average scores on the Stanford Achievement Test given at the end of the year. They also were given more praise and less criticism than the lows by the teachers. The direction of difference follows this pattern in all four classes for every variable in Table 2 except for the total correct answers, where the group means were equal in one class. Thus the Class $\times$ Expectancy interactions affected the degree but not the direction of expectation effects.

Sex effects also appeared, with boys producing more correct answers and receiving more criticism than girls. The other, nonsignificant differences in favor of boys are consistent with the finding noted above that boys tend to have more interactions with the teacher than girls. A Sex $\times$ Expectancy Group interaction occurs for the measure of total criticism which is similar and related to the one reported for behavioral criticism in Table 1. For the boys in the low group, teacher criticism was present in 32.50% of their dyadic contacts with the teacher. The corresponding figure for the high boys is 13.25%; for the low girls, 16.17%; and for the high girls, 8.25%.

In summary, the data of Table 2 show that teacher expectancy consistently predicts objective measures of classroom performance, objective achievement test scores, and rates of teacher praise and criticism. Hypotheses about the role of expectation effects in producing these relationships cannot be evaluated from the data in Table 2, however, since the type and direction of causal mechanisms at work remain unknown.

Group differences on variables interpretable as indexes of teacher expectation effects are presented in Table 3. Significant group differences on these measures suggest that the teachers were systematically, although not necessarily deliberately or consciously, treating one group more favorably than the other. The first two measures concern provision of response opportunities to the children, and may be considered in combination with the data previously discussed in Table 1. Since the highs create more response oppor-

tunities for themselves than the lows, do the teachers compensate for this by calling on the lows more frequently? The data suggest only a slight tendency in this direction at best. The teachers definitely do not compensate by asking the lows more direct questions, since the mean on this variable for the lows is less than that for the highs, although not significantly. The mean for direct questions in the low group would have been increased if "discipline" questions had been included in their figures. These were very special questions which appeared only in the low group, but not with sufficient frequency to be analyzed as a separate variable. "Discipline" questions were direct questions which ostensibly asked for academic content ("What's the next word, John?"), but which were directed at children not paying attention. In these instances the teacher's questions appeared to function as control techniques rather than as response opportunities, and so they were not included in the totals for direct questions. If they had been included the results would have been an increase in the mean for direct questions in the low group, but this mean value would still be below that for the highs.

The one teacher measure which does suggest some compensation concerns the teacher's behavior in calling on children to answer open questions. When the number of times the child is called on is weighted by the number of times he raised his hand to seek a response opportunity, the resulting recognition rates showed a significant difference in favor of the lows. However, this difference seemed more due to the large difference in hand-raising rate between the two groups of children rather than to any systematic compensation efforts by the teachers. The recognition rates in Table 3 are not adjusted for the fact that more highs than lows were likely to be raising their hands seeking an opportunity to answer a given question, so that a single response opportunity had less effect on the recognition rates of the highs than on those of lows. The rates may be adjusted by treating the highs and the lows as groups and discounting hand raising by other members of the group when one member of the group is called on. When the hand-raising totals are reduced in this manner, the resultant recognition rates still favor the lows, although the difference no longer approaches statistical significance.

In summary, the data on quantity of contacts in Table 1 and Table 3 are neutral with regard to expectation effects. The highs initiate more work-related contacts and create more response opportunities for themselves than do the lows, but there is no unequivocal evidence to suggest that the teachers are systematically either exaggerating or compensating for these differences among the children.

The data for the last five variables in Table 3 comprise the major findings of the study, since they provide direct evidence that the teachers' differential expectations for performance were being communicated in their classroom behavior. The measures involved are all concerned with the teachers' reactions to the children's attempts to answer questions and read in the read-

ing group. All are percentage or ratio measures which take into account absolute differences in the frequencies of the various behaviors involved so as to enable a direct comparison to be made between the teachers' behavior toward the two groups when faced with equivalent situations. The data show that the teachers consistently favored the highs over the lows in demanding and reinforcing quality performance. Despite the fact that the highs gave more correct answers and fewer incorrect answers than did the lows, they were more frequently praised when correct and less frequently criticized when incorrect or unable to respond. Furthermore, the teachers were more persistent in eliciting responses from the highs than they were with the lows. When the highs responded incorrectly or were unable to respond, the teachers were more likely to provide a second response opportunity by repeating or rephrasing the question or giving a clue than they were in similar situations with the lows. Conversely, they were more likely to supply the answer or call on another child when reacting to the lows than the highs. This group difference was observed both for difficulties in answering questions and for problems in reading during reading group. Finally, the teachers failed to give any feedback whatever only 3.33% of the time when reacting to highs, while the corresponding figure for lows is 14.75%, a highly significant difference.

Group differences in the direction of expectancy effects occur for all four classes on three variables; small reversals occur in the measure of criticism following wrong responses in one class and in the measure regarding teachers' reactions to reading problems in another class. These are the only measures for which the Class × Expectancy interaction is significant.

Significant sex effects also appear in Table 3, as they have previously. These show that boys receive more direct questions from the teacher than girls and that they are praised more frequently when giving correct answers. The difference on direct questions fits in with the general finding that boys tend to have more interactions of all kinds with the teachers than girls. The data concerning praise are more surprising, in view of the preponderance of criticism toward boys noted earlier. Taken together, the data on teacher praise and criticism in Tables 1, 2, and 3 suggest that the teachers are generally more evaluative in responding to boys and more objective in responding to girls. Boys are praised more often after correct responses and criticized more often after incorrect responses or failures to respond, although the latter difference is not statistically significant. The general preponderance of critical comments toward boys noted earlier is apparently due to behavioral criticisms rather than to critical comments made during work-related interactions.

Discussion

The data of Tables 1 and 2, which show objective differences among the children related to their sex and achievement levels, are quite consistent with previous findings. The finding

that high-achieving students receive more teacher praise and support (deGroat & Thompson, 1949; Good, 1970; Hoehn, 1954) was confirmed in the present study. Hoehn's suggestion that the differences between high- and low-achieving students in the interaction with their teachers were in quality rather than quantity of interaction is also compatible with present findings. The finding that teachers have more disapproval contacts with boys than girls has also been frequently reported (Meyer & Thompson, 1956; Jackson & Lahaderne, 1966; Lippitt & Gold, 1959). Meyer and Thompson, (1956) also reported greater praise toward boys, as was found in the present study in work-related interactions. Taken together, the findings on sex differences in the present study may be summarized as follows: boys have more interactions with the teacher than girls and appear to be generally more salient in the teacher's perceptual field. Teachers direct more evaluative comments toward boys, both absolutely and relatively. The largest and most obvious absolute differences in evaluative comments occur with teacher criticism and disapproval, which are directed far more frequently at boys. However, much of this difference appears to come in the form of behavioral criticisms and disciplinary contacts rather than criticisms of academic performance in work-related contacts. The difference appears attributable to more frequent disruptive behavior among boys which brings criticism upon themselves rather than to a consistent teacher set or bias toward being more critical toward boys than girls in equivalent situations. The latter statement agrees closely with the conclusion of Davis and Slobodian (1967), who studied teacher provision of response opportunities and evaluation of children's performance in reading groups.

While sex differences are attributable to objective differences in the classroom behavior of the children, the data in Table 3 show that differences related to teacher expectancy are only partly attributable to the children themselves. When the latter differences are statistically controlled through the use of percentage measures, it is seen that the teachers systematically discriminate in favor of the highs over the lows in demanding and reinforcing quality performance. Teachers do, in fact, communicate differential performance expectations to different children through their classroom behavior, and the nature of this differential treatment is such as to encourage the children to begin to respond in ways which would confirm teacher expectancies. In short, the data confirm the hypothesis that teachers' expectations function as self-fulfilling prophecies, and they indicate some of the intervening behavioral mechanisms involved in the process. Despite large differences in the frequencies of the various behaviors observed in the four classrooms, expectancy effects were consistent across the four teachers (two of the teachers favored the highs on four of the last five measures in Table 3, while the other two favored the highs on all five measures).

Although the direction of difference in treatment of highs and lows was constant across teachers, there were

observable differences in degree. In particular one teacher stood out as extreme in this regard, while one other showed relatively small differences, even though the direction of difference was constant. It is of interest that the latter teacher who showed the least discrimination between highs and lows was the teacher who did not group the children by achievement in her classroom seating pattern. It is also worthy of note that although the teachers' expectations were highly related to the children's achievement test scores within classes, the achievement scores are not so closely related to the previous readiness and achievement data which were used as the basis of tracking into classrooms. That is, the class achievement of some classes was higher than expected, while that of others was lower. While not enough classes were included to allow a statistical test, the data suggest that the achievement levels of the classes were related to the teachers' performance demands and expectations.

While this research has demonstrated the applicability of classroom interaction analysis methods to the study of the communication of teacher expectations and has yielded data which are consistent and interpretable as far as they go, it dealt with only a few of the events intervening between the formation of teacher expectations and the initiation of reciprocal behavior by the children. Several related studies are needed to complete the picture. For instance, if differential teacher treatment leads to differential reciprocal behavior by the children, the classroom behavior of highs and lows should become progressively more differentiated as the school year progresses. Hand raising, child initiation of work-related interactions, and other indexes of attempts by the child to seek response opportunities or academic interactions should show this kind of progressive differentiation between groups. Another set of questions has to do with intervention attempts. Can teachers be made aware of their discriminatory classroom behavior? Can they learn to compensate, not only for their own differential expectations but also for objective differences in the classroom behavior of the children? Will experimentally induced expectations produce the same differences in classroom behavior as expectations formed "naturally" by the teachers themselves? These and related questions will be taken up in future research.

Additional indexes of the ways in which teachers discriminate in their classroom behavior are also needed to add to our understanding of the processes involved and to increase the effectiveness of teacher education and classroom intervention in preventing or reducing the problem. Anecdotal observations taken during the present research suggest that other useful indexes of teacher communication of differential performance expectations may be possible. Possibilities being presently explored include differences in the type of feedback given to the children (inquiry into the processes underlying the response product rather than simple negation or provision of the right answer) and differential enforcement of teacher expectations (discouraging initiative in some children by

doing things for them while requiring other children to do the same things themselves). Teachers are frequently unaware of the subtle differences in their behavior in such situations, yet it is in such situations that teachers systematically communicate differential expectations to different students. Although subtle, such teacher behavior is observable and measurable, and therefore at least potentially subject to modification and control.

Summary

The processes by which teachers communicate differential performance expectations to different children were investigated through observational study of dyadic contacts between teachers and individual students in four first-grade classrooms. Differential teacher expectations for different children were associated with a variety of interaction measures, although many of these relationships are attributable to objective differences in the behavior of the children. However, other differential teacher behavior was observed which is not attributable to objective differences among the children and which is consistent with the hypothesis that differential teacher expectations function as self-fulfilling prophecies. The teachers demanded better performance from those children for whom they had higher expectations and were more likely to praise such performance when it was elicited. In contrast, they were more likely to accept poor performance from students for whom they held low expectations and were less likely to praise good performance from these students when it occurred, even though it occurred less frequently. The findings are interpreted as supportive of the hypotheses of Rosenthal and Jacobson concerning teacher-expectation effects and as indicative of the behavioral mechanisms involved when teacher expectations function as self-fulfilling prophecies.

References

Barber, T. X., and Silver, M. J. Fact, fiction, and the experimenter bias effect. *Psychological Bulletin Monograph,* 1968, **70** (6, Pt. 2).

Davis, O. L., Jr., and Slobodian, J. J. Teacher behavior toward boys and girls during first grade reading instruction. *American Educational Research Journal,* 1967, **4**, 261–269.

deGroat, A. F., and Thompson, G. G. A study of the distribution of teacher approval and disapproval among sixth grade pupils. *Journal of Experimental Education,* 1949, **18**, 57–75.

Good, T. L. Which pupils do teachers call on? *Elementary School Journal,* 1970, **70**, 190–198.

Good, T. L., and Brophy, J. Analyzing classroom interaction: A more powerful alternative. *Educational Technology,* 1971, in press.

Hoehn, A. J. A study of social status differentiation in the classroom behavior of nineteen third grade teachers. *Journal of Social Psychology,* 1954, **39**, 269–292.

Jackson, P. W., and Lahaderne, H. M. Inequalities of teacher-pupil contacts. Expanded version of a paper delivered at the meeting of the American Psychological Association, New York City, September 1966.

Lippitt, R., and Gold, M. Classroom social structure as a mental health problem. *Journal of Social Issues,* 1959, **15**, 40–49.

Meyer, W. J., and Thompson, G. G. Sex differences in the distribution of teacher ap-

proval and disapproval among sixth-grade children. *Journal of Educational Psychology,* 1956, **47**, 385–396.

Rosenthal, R., and Jacobson, L. *Pygmalion in the classroom: Teacher expectation and pupil's intellectual development.* New York: Holt, Rinehart & Winston, 1968.

Snow, R. E. Unfinished pygmalion. *Contemporary Psychology,* 1969, **14**, 197–199.

Thorndike, R. L. Review of R. Rosenthal and L. Jacobson, "Pygmalion in the classroom." *American Educational Research Journal,* 1968, **5**, 708–711.

10.2 *The American Teacher: A Tenative Psychological Description*

LOUIS S. LEVINE

Reprinted from *Psychology in the Schools,* 1970, **6,** 245–252, with permission of the author and the Clinical Psychology Publishing Company, Inc. The late Louis S. Levine was professor of psychology at San Francisco State College and was active as a consultant on research projects concerned with mental health and with the teaching of democratic values in the schools.

This article consists of an overview of a number of studies dealing with various aspects of teacher personality, with particular reference to the impact that recent changes in the social scene have had on the kind of person who is attracted to teaching and what their aspirations are.

The American teacher and the American school have been portrayed in many different ways. Teachers have been described as saintly and as sadistic, involved and indifferent, competent and incompetent, wise and stupid, resourceful and unimaginative. They have been characterized as knowing how to teach and what to teach and as being effective on both counts. Some have said the teacher knows only how to teach but does not know his subject matter. The severest critics of teachers maintain that the teacher neither knows how to teach nor does he know the substance of his subject. The evaluation of teachers is frequently linked with the evaluation of the schools and the society as a system. Thus, the range of observations pertaining to the system in which the teacher is a functioning element is also of interest to us here. Stereotypes of the schools have ranged from the "blackboard jungle" to the suburban educational sanctuaries.[1]

In any efforts to characterize American teachers and the schools, two salient points must be noted at the outset. The first point is that the wide diversity that exists among the psychological characteristics of teachers in the public schools is a partial function of their number. Presently there are over 1,250,000 elementary and secondary school teachers in the United States. The second point is that the schools are a reflection of the total society and its wishes with respect to what it wants taught and whom it wants to teach.

A bibliography compiled as background preparation for this paper which covers the span of years from 1957 to 1967 does indicate that the intellectual, personality, and motivational characteristics of the teacher have been of considerable interest to researchers. It does not, of course, follow that knowledge is necessarily directly proportional to the researcher's enthusiasms in choosing an area of inquiry. As a careful analysis of the total literature pertaining to the intellectual, personality and motivational characteristics of teachers reveals: There are few generalizations which can be drawn from the studies which have been conducted. Before offering the several tentative generalizations which appear to follow from the data presently available, several issues which complicate the interpretation of the available data must be mentioned.

The interlocking character of social process and individual behavior is such that the specific career choices that individuals make is determined by a host of economic and social forces operative at that point in time when the career decisions are made. The oppor-

[1] A summary of the conflicting views of teachers and schools is presented by M. A. Farber in *The New York Times,* January 28, 1968. In this summary he presents the view of Robert J. Havighurst, Visiting Professor at Fordham University, who called for a "moratorium on purely negative criticism" of the public schools. Havighurst's defense of the schools included an attack on those whom he classified as "non-responsible" authors. These individuals and their writings include: Friedenberg, E. Z. *Coming of Age in America: Growth and Acquiescence.* New York: Random House, 1965; Goodman, P. *Compulsory Mis-education and the Community of Scholars.* New York: Vintage Books, 1964; Holt, J. *How Children Fail.* New York: Dell Publishing Company, 1964; Kohl, H. *36 Children.* New York: New American Library, 1967; Kozol, J. *Death at an Early Age.* Boston: Houghton Mifflin Company, 1967.

tunities that exist for creative, intelligent, talented and poor young people are different in times of affluence, depression, war and peace. The behavior of the gatekeepers who determine admission policies at college, university, and occupational points is also a significant factor which influences the psychological and social characteristics of those who are permitted entry into given occupations and those who are closed out. The expectations and aspirations as well as the realities and perception of the realities depend not only on the life history of the individual and his transactions with the society, but also on the length of time in which his family has been within the culture and where they are relative to the mainstream of the society. The well-known phenomenon of the child of the immigrant entering teacher training as one of the fields offering excellent opportunities for upward social mobility is but a reflection of this latter point. Teacher-training institutions which twenty years ago catered to exclusively or predominantly Caucasian populations, offering the rationale that their Negro graduates found difficulty in being placed, now vie with one another for increasing the percentage of Negro students in their schools.

Individual studies relating to teacher characteristics require interpretation within the temopral and social context and in terms of how the social processes are manifest within the given geographic region and in the specific academic institution. In most of the studies undertaken that pertain to the characteristics of teacher trainees or teachers, there has been minimal attention given to effective descriptions of the individuals, their backgrounds, the institutional character of the training institutions or the schools in which they teach, or to the general play of the social forces relevant to career choices. Further, the effort to compare fragmented research studies is made difficult by the diversity in method, populations, procedures, measurement techniques, and absence of relevant social and psychological data. For example, in many of the available studies the typical research approach is to compare a group of students who may be enrolled in teacher training to a "comparable" group of students in another major. Personality differences have been frequently reported through the utilization of noting the statistically significant scale scores on personality inventories such as the Edwards Personal Preference Test, the California Personality Inventory or the Minnesota Multiphasic Inventory. When the question is whether such groups differ in interests, the Strong or the Kuder test results may be analyzed and a determination of differences between group means calculated. When this format is employed it is quite possible to attain statistically significant differences between groups on specific variables yet to have vast overlap between the characteristics of teachers and members of other professions with respect to personality and interest variables. The fact that statistical significance is often a partial function of the number of cases utilized in the study means that relatively small differences in mean scores between groups may occur when the sample is large. This

further means that for all practical purposes the likelihood of given individuals possessing the group characteristics may be exceedingly small and, for the on-the-line administrator, personnel officer, or college instructor, such findings may have little practical utility.

With these qualifications in mind, the findings relative to the intellectual, personality, and motivational factors can be summarized in the several generalizations that follow.

The Intellectual Capability of Teacher Trainees and Teachers

The general stereotype of the teacher trainee as being intellectually inferior to individuals who select majors other than education is not clearly supported by the available evidence. For example, in a study conducted in a municipal college in New York City during the late 1950s, over four thousand liberal arts and teacher-education oriented students were compared. Though slightly higher scores on the college entrance examinations were obtained for the intended liberal arts majors of both sexes, the differences were not statistically significant.[4]

Different colleges, universities, and teacher-training institutions vary with respect to the year of entry at which students interested in pursuing a teaching profession begin their academic and practical training. For example, at San Francisco State College—one of the largest teacher-training institutions in California—students traditionally begin their teacher-training programs during either the first or second year of their *upper division* work. In a study of all students who entered teacher training during the spring semester of 1957 and who were compared to the national norms on the ACT, no consequential differences were found in either verbal, quantitative, or total scores for men, nor verbal scores for the women. A small group difference was noted in the quantitative abilities of women education students (Levine, 1960).

The issue of entering teacher training is different from the issues of whether the individual completes teacher training or whether he enters and remains in the field of teaching as a professional career. In one of the most comprehensive studies undertaken, the careers of individuals who had taken a common battery of aptitude tests in the Air Force in 1943 were followed for a period of more than a decade. It was found that of the individuals who had been classroom teachers and college teachers, those who had demonstrated higher abilities in 1943 on tests of reading comprehension, arithmetic reasoning, and mathematics were more likely to have left the field of teaching (Thorndike & Hagen, 1961).

In view of the fact that intellectual ability, no matter how measured, contributes only approximately one-fourth

[4] Mitzel, H. E., and Dubnick, L. Relative scholastic ability of prospective teachers. *Journal of Teacher Education,* March, 1961, *12,* 78–80. The same authors published a rather complete review of data obtained from the nationwide Selective Service screening in 1950 to 1952, from which they were able to compare the norms of teachers college students with those from liberal arts schools in terms of their scholastic aptitude and test results. The authors found a high degree of variation and concluded that the charge of inferior academic ability of teachers who entered teacher training during this period was unsubstantiated.

of the variance to any measure of attainment—academic, or otherwise—the definitive issues relating to the psychological characteristics of teachers would appear to be their personality and motivation.

The Personality of the Teacher

Because the teacher is a central factor in creating the conditions conducive to classroom learning, the assumption is made too readily that the one single set of characteristics descriptive of the personality of the good teacher will carry over to all teachers at all levels. The findings do not support this contention. First, the psychological characteristics of men and women who enter and complete teacher training are quite different. Also, there are differences between elementary and secondary school teachers and between teachers and administrators with respect to personality characteristics, and there are personality differences among those who enter the various schools that provide teacher training.

The need to specify the context in which the teacher is functioning is demonstrated in the findings of a study in which personality data were obtained from practicing teachers in the Chicago area, teacher trainees at a mid-west state university, a southern Negro college, and a private urban teachers college. The authors of this study put forth three hypotheses, all of which were corroborated. These hypotheses were: (*a*) Trainees who chose to enter a multi-purpose institution will demonstrate personality patterns that are responsive to the press of the institution rather than to the press of the profession. Conversely, teacher trainees who enter a teachers college will display personality patterns which resemble those of the practicing professionals. (*b*) Teaching experience tends to erase particular need structures that were responsive to the press of the training institution and produces a pattern that corresponds to teaching groups regardless of their academic background. This pattern, the authors suggest, is characterized by being highly deferential, placing a premium on order and endurance and of low heterosexuality (as judged in terms of prevalent interest patterns), and of high dominance and the need to perform. (*c*) For a given school, the more nearly the teachers approximate the typical teacher-personality patterns the less likely they are to feel satisfied, effective, and confident in the ability of their administrative officials, and the more likely the administration is to regard them as effective (Guba, 1959).

Findings of the study indicated above are of particular interest because they point out that personality characteristics of individuals who choose to enter certain colleges differ, and that the college itself may have a differential effect on the personality characteristices of the student. Thus, the problem of specifying psychological characteristics of personality of teachers independent of the context either in which they are studying or working is somewhat hazardous. Yet, the counter expectation would seem to follow (also on the basis of the Guba study) that time in the profession itself seems to erase the influences of earlier experience and the

nature of the academic institution. And to an extent this is true—keeping in mind the qualifications previously noted—the teachers who demonstrate deferential responses to their administrators and to parents are apt to be judged favorably by them. Efficiency and physical endurance undoubtedly are required in dealing with large numbers of students over six or so consecutive hours of the day and usually with little opportunity for the usual aesthetic or physical amenities, as Friedenberg has noted (see Footnote 3). The finding that teachers, in general, score low on the measures of heterosexuality must be very clearly understood as a comment on teacher interest, rather than on their actual sexuality. A man who is interested in literature, painting, people and music will tend to receive a low heterosexual score. These scores are really matters of interest and taste rather than of sexuality.

Another theme that runs through a number of studies is the tendency, particularly of the male teacher, toward an authoritarian and rigid personality. In the standardization population of the California Personality Inventory, it was found that next to military officers, teachers, as a group, were the least flexible of all of the occupations represented in the test standardization population. The study conducted at San Francisco State in the late 1950s that pertained to individuals who entered the teacher training program, found that there were statistically significant differences between individuals entering teacher training and the normative population for the California Personality Inventory. The male teacher trainees were more aggressive, persuasive, verbally fluent, outgoing, enthusiastic, spontaneous, competitive, energetic and self-centered than the population of all college men. Also, they were more inclined to be dogmatic, undercontrolled, impulsive, opinionated, rebellious, undependable, assertive and more concerned with self-gain than the total population of all college males. The men entering teacher training were also noted as more appreciative, patient, helpful, and gentle than the general population of all college men (Levine, 1960).

In the same study, the women teacher trainees at point of entry into their professional teacher programs, when compared to all college women, were inclined to be more self-confident, self-assured, more verbally fluent, cooperative, helpful and diligent. But they were also more conventional and less resourceful, more restricted in their general outlook and range of interest, and more in need of supervision and direction. In this specific study, when the men and women entering teacher training were compared, the men were scored as more impulsive, aggressive, assertive, and egocentric than the women and also less mature, responsible, and self-controlled. This comparison is of special significance. The tendency to fill administrative openings in public schools with men, and that they are preferred over women as administrators by other administrators and by boards of education, does not appear to operate in the best interest of the schools or the needs of the children. The inequity to women of considerable competence and talent in this situation

is, of course, discriminatory in practice, and is a subject that goes beyond the scope of this present paper.

Teaching and Motivation

The point has been made earlier that the motives which influence specific individuals to seek entry and to remain in teacher training and in the teaching profession vary with the social and economic conditions prevalent during specific periods of time. The study conducted at the University of Montana during the 1964–65 term is considered representative of the studies during this period that utilize data based upon student questionnaires and student interviews (Hood, 1965). In this study, the 226 university students who had selected teaching as a career indicated that they believed teachers perform a valuable service to society and that teaching affords them the opportunity to work with young people. This statement of interest by teacher trainees in working with young people holds up through time and has been reported in various studies over the last 30 years. The consistency of this response, and the fact that it is part of a role or social expectation, leads one to question its significance since in several studies in which individuals have attributed their interest in teaching to their desire to work with young children they had relatively minimal exposure to young children. However, returning to the statements made by the University of Montana students in the 1964–65 period, they also noted that the teaching profession affords a springboard or avenue of entry into other fields. The impression of the present writer is that in many instances the decision to enter teaching is arrived at by default rather than by design. Confronted with the necessity of making a specific occupational choice, the student about to enter college or the student in college who is faced with deciding on his major is influenced by the availability of opportunity and by the number of alternatives that the career choice affords. Thus, business education may be selected as a major by an individual who is not totally certain whether he actually wants a business career or is uncertain as to the specific aspect of the commercial world that appeals to him. The well-known phenomenon of the young woman who believes that elementary school training will help her in raising her own children represents the *motivation* of many women who make the career decision to enter the profession of teaching. Here the major interest may be to marry and raise a family and the choice of the professional career is probably viewed as congruent with her primary aspiration.

In the Montana study cited above, the prospective teachers stated that one of the disadvantages they perceived in the career of teaching was that the personal freedom of teachers is restricted in certain communities. Further, they stated that teachers' salaries are less than those paid in many other professions requiring the same amount of training, they might have to present subject material that they were unqualified to teach, and the expected work load would be excessive.

The unfavorable views which teacher trainees hold toward teaching condi-

tions are not unrealistic as the available literature indicates. In a summary of the relevant research on this issue,, the following factors were consistently stated by individuals who had taught for a year and left the teaching profession: Income was inadequate; teaching loads were excessive; teachers were assigned duties that ranged far beyond the classroom or teaching activities; supervisory assistance was inadequate or not available; and the assignments given to first-year teachers were felt by them to be inappropriate.

In addition to the five explanations above, the list of frequently cited reasons for leaving the teaching profession included inadequate preparation, inadequate facilities, and a lack of opportunity to develop new ideas. Some of the teacher drop-outs complained of routine clerical duties and poor faculty relationships as their major reasons for leaving the field of teaching (Nelson & Thompson, 1963).

The perceptions of students entering teacher training, particularly the disadvantages about which they are concerned and the reasons offered by first-year teachers who drop out of the teaching profession, are congruent with the results of a study reported in 1963 which covered a time span of 33 years. Of those who responded, 14% had left public-school employment. These teachers and administrators stated that their major reasons for leaving were inadequate salaries, that they experienced a lack of satisfaction in the teaching profession, and that they felt the respect and status which they deserved within their communities was not forthcoming (Willey, 1963).

Again it must be noted that the above characterizations of teachers can only be accepted tentatively because of the finding that in most of the studies there is an almost complete overlap of the intellectual, personality, and motivational attributes between the teachergroups and other occupations.

It is this writer's opinion that current teacher-training populations may have a higher incidence of youthful activists and idealists than was the case even five years ago. These impressions have been reported to the writer from several schools and could be due to the larger number of youthful militants in all school programs than some years ago. There is also the possibility that increased numbers of dedicated youth want to contribute to creating a better world and are deliberately entering teaching in order to work within the urban ghetto with the so-called disadvantaged. Another question that this writer has not been able to substantiate empirically is whether there is now a higher incidence of black students in teacher-training programs than was the case a decade ago. The assumption that this is so seems defensible on several counts. One is the desire of many institutions—teacher training and otherwise—to enlist and, in fact, to compete for Negroes. Whether the recruitment efforts will substantially increase the number of Negro students in teacher-training programs, is, of course, conjectural at this point. At San Francisco State College, for example, where the number of individuals who are enrolled in teacher-training constitute approximately 34% of the total enrollment of the institution, the total Ne-

gro student population is just slightly above 4%. There is no evidence to indicate that a higher incidence of this relatively small percentage of Negro students in the school enter the field of teacher training as compared to any other major. A year ago the University of California at Berkeley, with some 27,000 enrolled students, had a Negro student population of less than 1%. Since then, though special efforts have been made to increase the undergraduate enrollment, the ultimate outcome of these efforts, as concerns people entering the teaching field, is as yet uncertain.

Pursuing a speculative theme, here, we can anticipate that a number of changes may be occurring in the types of individuals who will seek admission to teacher-training programs and who will persist in teaching careers. We note, of course, that as yet in this paper we have not discussed the issue of the intellectual, personality, and motivational characteristics that are correlated with the various measures of teaching effectiveness. The difficulty of identifying a clear-cut criterion of teacher effectiveness is well-known, as is the fact that such determinations when made by administrators, colleagues, and students will often vary markedly. In general, the studies which have utilized various techniques to identify good teachers follow the results obtained by Ryans (1960). He generalized that the good teacher shows superior intellectual ability and above-average adaptation to his own school experience. According to Ryans, a good teacher is relatively well adjusted emotionally, demonstrates favorable attitudes toward his pupils, and enjoys contact with them. He is generous in his appraisal of the behavior and motives of other persons, and has a higher than average rate of participation in social and community activities. The good teacher is apt to have a higher order of interest in reading, music, and painting than teachers who are judged less effective, and to show strong social service interests.

In one sense, the statement describing the good teacher could equally well describe the "good person" in a democratic society. The issue is whether the qualities that presently characterize individuals who enter and remain in teaching, or even those which characterize the "good" teacher, are necessarily those which should be employed in the selection of teachers or be viewed by those in charge of their training as appropriate to the future teaching situation. Consider, for example, the pressures toward school decentralization in the major urban centers. If schools are decentralized, then the competencies involved in working much more closely with parents—particularly parents of children from minority groups—will require greater emphasis. It is a commonplace observation that some individuals who are excellent at working with young children are not particularly comfortable in their interpersonal relationships with adults. To the extent this is so (if it is) what types of in-service and pre-service training can be brought to bear to extend the range of teacher competencies? This problem represents a compelling challenge to those concerned with teacher training. To what extent will all teachers

—those in suburbia as well as those in the inner cities—have to become more knowledgeable and more informed about the history and contribution the Negro American has made to this society? The basic elements of prejudice, whether they arise from neurosis or from ignorance, are issues that the total society will not be able to avoid in the years of racial crisis ahead. Whether the educational enterprise will be able to contribute to breaking the equation that difference equals danger, of course, constitutes one of the major questions of our time. Its answer will depend, in part, on the degree to which the teachers of the future will be able to act appropriately and be able to face difference without fear (Levine, 1966).

The transition of the teacher's role from that of primary informational source to that of counselor, in the best sense of the term, which will accompany the technology and curriculum modifications that lie ahead will require marked changes in the teacher's ability to deal with students as individuals; the special competencies required of adults who contribute to the full development of young people will have to be given more attention in pre- and in-service teacher training.

Those presently concerned with teacher training undoubtedly see the full range of the human characteristics mentioned above reflected in the persons who arrive at their classrooms. Undoubtedly, most college professors place a high value on facilitating the learning of their particular subject matter. Unfortunately, we don't know, for example, whether the subject matter that is presented in most college programs is, in fact, relevant to the development of children or young people; nor do we know whether or not the acquisition of knowledge, as such, in any way contributes to the classroom behavior of the teacher. Yet, we do know that to the typical college professor, *his* is the important course. The curriculum "imperative" is automatically evoked when new courses or programs are proposed to faculty committees. The descriptions of the *status quo* courses presented eloquently by their defenders would cause the students who have taken these same courses to wonder with amazement how the unusual merit of these classes had managed to evade them.

Until adequate longitudinal studies of teachers are available, the problem of adequately characterizing the teacher, his classroom, his school, and the community and the implications of such findings, for society in general and teacher-training specifically, can only be approached most tentatively.

References

Getzels, J. W., Guba, E. G., and Jackson, P. W. Occupational choice and the teaching career. *Educational Research Bulletin,* 1959, **38**, 1–12.

Hood, C. E. Why 226 university students selected teaching as a career. *Clearinghouse,* 1965, **40**, 228–231.

Levine, L. S. *Students Entering and Completing Teacher Training,* a monograph. San Francisco State College, 1960.

Levine, L. S. Imposed social position: Assessment and curriculum implications. *The Bulletin, NASSP,* 1966, **50**, 44–59.

Nelson, R. H., and Thompson, M. L. Why

teachers quit: Factors influencing teachers to leave their classroom after the first year. *Clearinghouse,* 1963, **37**, 467–472.

Ryans, D. G. Profile of the good teacher. *School and Society,* 1960, *88,* 424.

Thorndike, R. L., and Hagen, E. Men teachers and ex-teachers: Some attitudes and traits. *Teachers College Record,* 1961, **62**, 302–316.

Willey, R. L. A behind the scenes look at some former teachers. *School and Community,* 1963, **49**, 15–41.

10.3 *Authoritarianism, Independence, and Child-Centered Practices in Education: A Study of Values*

HENRY CLAY LINDGREN

Reprinted from *Psychological Reports,* 1962, **10**, 747–750, with permission of the Southern Universities Press, Inc.

Common sense would suggest that the authoritarian person, with his harsh and drastic methods, is likely to be more independent than most people. Research with various measures of authoritarianism and independence does not confirm this, but finds the opposite: it is the highly authoritarian person who is likely to be the most conforming and least independent. Where do learner-centered teaching methods fit in? Common sense might say that they would be most strongly endorsed by democratically minded individuals, but at the same time they sound more suggestive of softness and permissiveness, rather than independence. What the present survey shows is that child-centered, democratic modes of teaching tend to be favored by persons who have more education, who think independently, and who score low on a scale of authoritarianism. The results, which have been confirmed by crosscultural studies in Brazil and in Lebanon, can be interpreted rather simply: 1. permissiveness and child-centeredness tend to be counternorm, because adults are generally more concerned about controlling children than understanding them; and 2. any individual who supports a policy of permissiveness and understanding in dealing with children must of necessity be independent in order to resist normative pressures to abandon such ideas. There is no reason to expect that teachers would be any different from the nonteaching public in this respect, namely, that those teachers who are the least authoritarian and most supportive of learner-centered methods are also likely to be the most independent in thought and behavior.

Although common sense suggests that child-centered attitudes and practices in education (as contrasted with traditional attitudes and practices) would be more consistent with equalitarian than with authoritarian attitudes, there appears to be little research that directly confirms such a supposition, particularly with respect to nonteaching personnel. Data regarding such a relationship would be of interest and value especially in the United States, where the determination of basic educational policy is the responsibility of governing boards of laymen, who have in turn been elected or appointed by other laymen. What layment think or feel about education is thus of considerable importance in determining the direction of education in this country.

This study was undertaken in order to determine whether the acceptance by nonacademic personnel of child-centered practices in education would be negatively correlated with authoritarian attitudes. The author's access to a nonacademic sample also provided him with the opportunity for studying other variables that showed promise of being related to child-centeredness. Previous research with college students (Lindgren, 1961) had suggested the possibility of a positive relationship between child-centeredness and attitudes characterized by independence, but the relationship between the two variables was at a low level of significance for the males in the sample, possibly because of restriction in range for one of the variables being measured. Furthermore, the level of child-centeredness in the group being tested was so high as to suggest a substantial degree of relationship between educational level and child-centeredness.

Method

The problem of locating a nonacademic population for purposes of research is often difficult for behavioral scientists who function in an academic setting. In the present study, the problem was solved by taking advantage of the availability in San Francisco of two coin-operated Laundromats, one located in an area of middle-class apartment houses and upper-middle-class homes, and the other located in a "fringe" area of working men's hotels, transient rooms, and lower-middle- to lower-class apartments and homes. Laundromat customers seemed particularly desirable for a study using questionnaires, not only because they represent more of a cross-section of the general population than do college students, but also because they have to spend a half hour to an hour waiting for their laundry to be processed by the machines and, having time on their hands, are likely to be receptive to the idea of cooperating in a psychological study.

In the present study, the author attempted to maximize participation and cooperation by paying the cost ($.30) of washing and drying 1 9-lb. load of laundry for each respondent. Not more than a tenth of the customers refused to participate. In some instances they refused because they had planned to shop while their laundry was being processed; in other instances, particularly in the fringe area, refusal appeared to be based on the customer's inability to read.

The degree of acceptance of child-centered policies and practices in education was measured by a 30-item version of a 50-item questionnaire that had been developed and used in connection with a study by Lindgren and Patton (1958) and had been used as well in the study correlating independence with child-centeredness referred to above (Lindgren, 1961). The 50-item questionnaire had a corrected split-half reliability of .82. It was shortened (in order to keep it within the confines of one page and thus reduce the possibility of resistance and fatigue) by eliminating the 20 items having the lowest correlation with the total score. The resulting 30-item instrument had a corrected split-half reliability of .64. The items comprising the questionnaire touch upon three areas of educational policy: (a) the desirability of understanding the behavior of students, particularly in terms of its psychological causation; (b) the desirability of the teacher's using authoritarian methods as means of controlling the behavior of students; and (c) subject-matter-centeredness vs. learner- or child-centeredness. Here are some samples of the kinds of items in each of these areas:

12. Most boys and girls who present extreme cases of "problem behavior" are doing the best they can to get along with other people. ("Understanding")

3. It is appropriate for teachers to require an additional assignment from a student who misbehaves in class. ("Control")

4. How a student feels about what he learns is as important as what he learns. ("Subject-matter-centeredness vs. child-centeredness")

Some items fall into more than one category:

15. Boys and girls in the elementary school should be promoted regardless of whether they have completed the work on their grade or not.

In addition to the questionnaire on attitudes toward education, respondents were asked to complete the following items: (a) a brief questionnaire covering years of education completed, sex, year of birth, and occupation and (b) a 22-item questionnaire composed of items found by Barron (1953) to discriminate at the .05 level or better between "independents" [those who did not conform to the incorrect majorities in the situational tests devised by Asch (1956)] and "yielders" (those who always conformed). This questionnaire has been cross-validated by Tuddenham (1959) who found that it discriminated between yielders and independents for men (but not for women). (c) They were also given a 7-item form of the F scale used by MacKinnon and Centers (1956) in a study conducted in Los Angeles of the interrelationship among age, occupational status, education, income, race, and authoritarian attitudes.

Results

The results of the study are presented in Table 1. They confirm the commonsense observation that child-centered attitudes regarding educational practices and policies are more consistent with democratic than with authoritarian beliefs. They also confirm the finding of the previous study (Lindgren, 1961), namely, that the degree of child-cen-

TABLE 1. Pearson Product-moment Correlations among Child-Centeredness in Education, Authoritarianism, Independence of Judgment, Years of Education, and Age

	Authoritarianism	Independence	Education	Age
	Men ($N = 81$)			
Child centeredness in education	—.28**	.36***	.19*	—.23*
Authoritarianism (F scale)		—.29**	—.31**	—.05
Independence of judgment (Barron's test)			.32**	—.08
	Women ($N = 69$)			
Child-centeredness in education	—.51***	.57***	.38***	—.35**
Authoritarianism (F scale)		—.59***	—.29**	.25*
Independence of judgment (Barron's test)			.34**	—.21*

*$p < .05$.
** $p < .01$.
***$p < .001$.

teredness is positively related to the degree of independence of judgment. In general, the more years of education completed, the greater the degree of child-centeredness. Like the other relationships reported in this study, this finding was even stronger for women than men. Age was negatively correlated with child-centeredness.

Other findings of general interest, although not directly related to the focus of the present study, are the significant and negative correlations between authoritarianism and independence and between authoritarianism and educational level. The correlation between educational level and independence is both positive and significant.

Summary

An analysis of the questionnaire responses of a sample of Laundromat customers showed signifificant and positive correlations among years of education completed and attitudes characterized by equalitarianism, independence, and acceptance of child-centered practices in education. Age was negatively correlated with these attitudes.

References

Asch, S. E. Studies of independence and submission to group pressure: 1. A minority of one against a unanimous majority. *Psychol. Monogr.,* 1956, **70**, No. 9 (Whole No. 416).

Barron, F. Some personality correlates of independence of judgment. *J. Pers.,* 1953, **21**, 287–297.

Lindgren, H. C. Correlates of attitudes toward child-centered practices in education. *Psychol. Rep.,* 1961, **9**, 440.

Lindgren, H. C., and Patton, G. M. Attitudes of high school and other teachers toward children and current educational methodology. *Calif. J. Educ. Res.,* 1958, **9**, 80–85.

MacKinnon, W. J., and Centers, R. Authoritarianism and urban stratification. *Amer. J. Sociol.,* 1956, **61**, 610–620.

Tuddenham, R. D. Correlates of yielding to a distorted group norm. *J. Pers.,* 1959, **27**, 272–284.

10.4 *Anxieties and Discontents in Teaching*

ROBERT H. SNOW

Reprinted from the *Phi Delta Kappan,* 1963, **44**, 318–321, with permission of the author and Phi Delta Kappa, Inc. Robert H. Snow is program administrator in the Adult Education Division, University College, Syracuse University.

A major source of identity for modern man is his work. We are to a large extent what we do, and it is through what we do that we discover who we are. Individuals in most professions receive continual feedback as to their success and failure, which not only helps them improve their performance, but also gives them information as to the extent to which their self-image is valid. Teaching, as Snow points out, generates little feedback and, as a consequence, leaves the teacher with a feeling of vagueness both as to the validity of his efforts as well as his self-image. Snow suggests that this can be corrected if teachers would recognize the problem and do something about it. They must develop ways of evaluating their performance, partly as a way of self-affirmation and partly as a way of getting information to communicate about their profession. Without such data, teachers run the risk of being the most misunderstood profession—misunderstood by themselves, as well as by others.

Today the teacher is being recognized as the key to school improvement, and commendable steps are being taken to help teachers operate more effectively. Instructional materials and equipment are being improved. Students are grouped to permit greater efficiency in teaching. Teachers are being relieved of some clerical duties. Salaries have been increased appreciably. Abundant opportunities for advanced professional training have been established.

Still, no broad-scale transformation of classroom practice has occurred. By and large, teachers across the nation still teach much as they have for the last quarter century. While notable changes have been occurring in medical practice, in merchandising, industrial production, and virtually every other field of human endeavor, the practice of teaching remains essentially unchanged.

Persons concerned with improving our schools and colleges are wondering why this is so. Financial limitations undoubtedly play a part, but do not tell the whole story. The many problems involved in educating all youth rather than only a selected few place tremendous burdens upon teachers. However, these problems are not insoluble, nor do they constitute the sole reason why progress is painfully slow. Explaining the failure to achieve excellence in teaching by citing the laziness, indifference, or stupidity of teachers is transparently foolish. Some causes lie deeper, and to find them we must examine the teacher's problems in greater detail.

A teacher is, by the very nature of his work, denied clear-cut, indisputable proof of his effectiveness. He has no dependable means of tracing the consequences of his teaching, of discerning the precise extent to which his efforts have helped students learn. Even when gains made by students during the period of exposure to a teacher's services can be measured, the teacher cannot be confident that such progress is directly attributable to his efforts. Other factors may have been responsible. He must face the possibility that his own contribution may have been negligible, or even negative.

Many of the teacher's perplexities and frustrations stem from his fundamental predicament. Because he has no satisfactory means of evaluating his performance by results produced, he must resort to supposition and inference. In teaching, he must choose among alternative courses of action somewhat blindly, trying to decide which course he can reasonably assume will be most effective. And having done so, he can gain no clear indication, through observation of results, of whether he is proceeding correctly. In effect, he is steering a ship in semi-darkness, with a compass he can scarcely see.

In this circumstance it is not surprising that teachers generally follow traditional practices and resist innovations. When it is impossible to demonstrate that one method of operation is superior to another, custom tends to prevail. Stereotyped methodology becomes perpetuated. How can one seek out a better mode of operation when "better" and "worse" appear to be indistinguishable?

This uncertainty may have profound

effect on the attitudes and conduct of the teacher. In subtle ways, he may be tempted to neglect teaching in favor of other activities, both within and outside the school, which will afford more positive assurance of his competency. He may be reluctant to examine indirect evidence which might shed light on the effectiveness of his teaching and serve as a guide to improvement. Finally, his equivocal position may lead him into routinized habits which are actually detrimental to good teaching.

Despite pious pronouncements about the nobility of his calling, the career teacher in America finds it difficult to attain a favorable self-concept through his efforts in the classroom. He lives in a dynamic, work-oriented society, where reputation is gained and a sense of personal worth is established largely through achievement, and in particular through achievement in one's vocation. Competency in work performed is a major criterion by which we judge ourselves and others, and the normal craving for a favorable self-image will rarely be fully satisfied unless vocational achievement confirms and justifies the image. The difficulty in finding evidence of vocational proficiency becomes a source of frustration for many teachers.

Self-Evaluation Based on Others' Judgments

Because the teacher is committed to an enterprise in which he cannot hope to establish objective proof of his competency, he must base his self-evaluation largely upon the implied judgments of other persons. In this respect, his position is not totally different from that of the physician, the lawyer, or any other professional worker. A craftsman who fabricates a product may know his success by examining the product. A salesman may judge his success by the amount of his commissions, the merchant or manufacturer by the profit on his books. In the main, professional work does not offer such possibilities for direct evaluation of performance by results. Accomplishments are vague and indeterminate. But in most occupations, indirect measures of achievement and various forms of recognition and reward serve as tokens of fulfillment and guides to future action.

Although the physician cannot be absolutely certain that he deserves credit for the recovery of his patient, he has gratifying confirmation of his expertness in the fact that patients continue to seek his services and are willing to pay for them. Although the lawyer may not be sure that agreements are consummated or conflicts resolved entirely through his intervention, he is sustained by the realization that his services are in demand, and he enjoys the deference accorded him as a member of an honored profession.

The teacher, similarly unsure that he has produced whatever learning has taken place, does not even enjoy the knowledge that his students came to him freely. For the teacher a repeat customer is a sign of failure, not success.

The prevailing cultural judgment of the teacher's worth, reflected in his financial rewards and in the patronizing manner so often displayed toward teachers, suggests that public endorsement of his work is less than enthusi-

astic. The immediate benficiaries of his services do not seek him out. Ordinarily he operates within a bureaucracy, public or private, and his "clients" suffer exposure to his services, in many instances with obvious reluctance, as they pursue broader objectives. Students pay little tribute to his work, often displaying the utmost indifference as he attempts to assist them. The fact that he has been singled out for employment within a school or college and is continued in his job offers slight gratification. He knows that many incompetents are also employed because scarcity of candidates makes it difficult to keep positions filled. Increases in salary imply no recognition of personal merit, because they are granted to all teachers in the system according to years of service. Promotion, if it comes, ordinarily takes him out of the teaching ranks and into administrative echelons. The very fact that he remains a teacher seems to imply that he is unworthy of promotion.

With little assurance that his classroom services are valued by others, and unable to refute adverse judgments with clear evidence of accomplishment through teaching, the teacher may find other pursuits more attractive. He may throw his energies into coaching an athletic team, staging dramatic performances, or engaging in other school-related activities where his services are more conspicuous and recognition is more easily achieved. The proliferation of team sports, public performances, and gala special events within our schools has been deplored in some circles as a distraction from serious study. We forget the contributions these activities make to the sponsoring teacher's ego.

Some teachers develop sidelines outside the profession, engaging in business ventures, officiating at sporting events, or holding public office, and not merely to supplement their incomes. Success in these undertakings may compensate for disappointments in the classroom. If substantial income can be derived from them, competency is further attested.

Publications the Badge of Professional Success

One suspects that the marked inclination of college professors to devote attention to matters outside the classroom is, in great part, similarly motivated. Publication is often cited as the route to academic andvancement. It may also be noted that a book or article in print stands as tangible evidence of accomplishment, more conspicuous testimony to the author's ability than he can hope to gain through classroom service. Even without financial inducements, engagement as a consultant gives comforting reassurance that one's services are in demand.

However, as the teacher's energies are diverted from the classroom, the essential teaching function of the school is weakened. Less attention is given to the improvement of instruction. Routine procedures become established. Practices of questionable value in teaching but demanding little ingenuity or resourcefulness on the part of the teacher are adopted, and the students are deprived of the assistance to which they are entitled.

Improved performance by any teacher requires some method of gauging relative success or failure so that corrective action may be taken. If a rifleman on the firing range should be blindfolded and denied information as to whether his shots were striking the target, his marksmanship would not improve.

No Improvement without Information

If teaching is to improve, there must be a continuous channeling back to the teacher of reliable information about the effectiveness of his efforts so that future teaching may be adjusted for better results.

The teacher meets many discouragements. The achievement of most students under his tutelage falls far short of his aspirations for them. His status within the cultural *milieu* is a modest one. His daily encounters often suggest that he is regarded as ineffectual. Under such circumstances, the teacher may well become defensive, insecure, and sensitive to criticism.

It is easy for him to conclude that all appraisals of his work are apt to be unfavorable. Hence, he may resist strenuously any systematic attempt to evaluate his performance as a teacher. The mere presence of an outside observer in his classroom becomes a threatening invasion of privacy. The rating of his performance by supervisors or by students may be regarded as humiliating.

When such attitudes take over, the teacher may deprive himself of valuable information which could heighten his effectiveness and even sustain him in a more gratifying self-image. Yet there are few signs of eagerness on the part of teachers for information about their teaching and many indications of active resistance. For example, pretesting of the students' knowledge at the beginning of a course will permit comparison with scores at the conclusion and provide some basis for measuring progress made during the period of study. In many subject areas, standardized tests are available for this purpose. Such tests will also supply the teacher with useful information for individualizing instruction, selecting topics for special emphasis, and avoiding needless repetition. However, pre-testing is extremely uncommon. Apparently most teachers prefer to assume that students begin in a state of total ignorance and that all learning which can be demonstrated at the conclusion has been acquired during the course.

Opinions of students regarding the way educational programs are conducted and the kinds of experience they find helpful can be another source of valuable information for the teacher. Yet even on advanced college levels student opinion is rarely solicited in a systematic manner. More commonly, elaborate safeguards are erected against the overt expression of student judgments. On the pretext that students are unqualified to form sound opinions, the teacher's "dignity" is preserved, and as a result useful insights are sacrificed. Once past the period of apprenticeship, a teacher seldom receives counsel from another adult who has observed that teacher in action for an extended period. It is difficult for a

teacher to make an accurate appraisal of his own performance. The perceptions of a sympathetic and knowledgeable colleague can be immensely helpful. Unfortunately, few teachers avail themselves of such assistance. The emotional overtones are too great.

Improvement in teaching often demands a break with tradition, a willingness to discard conventional practice in favor of innovations. The teacher's anxieties can seriously impede necessary experimentation. When confidence is shaky, the *absence* of reprimand offers some semblance of approval, some confirmation that one is not hopelessly astray. There is always the possibility that, if one attempts something different, results may be less favorable than at present. Therefore it appears unwise to alter existing procedures and jeopardize what little security has been achieved.

Emotional considerations often cause barriers to be erected between teacher and student. Many students are enrolled against their will. Because of legal compulsions, parental pressures, fear of economic and social consequences attendant upon withdrawal from school, students find themselves trapped in a relationship with teachers. Some experience a continuing succession of failures within the academic setting. Rebelling against a situation not of their choosing, they may display undiscriminating resentment toward many aspects of the school environment, including their teachers. They may be disorderly, uncooperative, reluctant to pursue assigned tasks, and apparently indifferent to learning. Such feelings may be shared by half the students enrolled in the later elementary grades and early high school years and by a considerable number who complete high school and enter college. To work effectively with these students, the teacher must demonstrate exceptional emotional stability.

Rigid Discipline Reassures the Teacher

In the face of many uncertainties, the teacher may feel a desperate need for exercising rigid control within the classroom to insure that at least the outward appearance of constructive effort is maintained. When students seem orderly and attentive, it is easier for the teacher to feel assured that he is teaching successfully. Deviations from accepted behavior patterns must be sternly suppressed, because they destroy this sense of confidence. Furthermore, evidences of sloth or recalcitrance are interpreted as personal provocations, because they suggest that the teacher's services are not appreciated.

Some teachers appear to live in dread of insurrection, obsessively concerned with preserving order in the classroom at all costs. Coercive measures predominate; reproaches fill the air. Inordinate amounts of time are spent enforcing minor regulations. The examination becomes a punitive device, chiefly intended to place students on the defensive rather than to measure achievement. The student regards the teacher not as one who guides and assists but as one who threatens and invokes penalties. A gulf widens between students and teacher. The class-

room becomes an arena of opposing forces rather than a laboratory for learning.

It is equally unfortunate when teachers, bedeviled by pressures and anxieties, become contemptuous of students. If students are viewed as unworthy of attention or incapable of learning, the cause is lost. A teacher can perform at his best only with sincere interest in the welfare of his students and hopefulness regarding their potentialities. When the teacher must shift responsibility for failure to students in order to preserve some remnants of self-esteem, the foundations of good teaching are undermined.

Certainly many dedicated teachers do focus their talents and energies upon their work with students, seeking no reward beyond the intrinsic satisfactions of a life of service. Yet it is sheer sentimentality to believe that teachers are a race apart, noble and selfless, transcending normal human appetites. The nation cannot build an adequate educational system upon a false assumption of this magnitude.

If professional service is expected of teachers, financial rewards must match those of other professions. If the teacher's role is central in education, this role should be recognized and other functions made subordinate to it. In our present organization, the teacher is lowest on the educational totem pole. Sometimes he is paid less than the janitor and usually less than the football coach. In an educational institution, business managers, counselors, administrative officers and other service personnel have one basic responsibility: to sustain and facilitate the work of the teacher. Our operating procedures often suggest that the teacher's function is less significant than many of these supportive efforts. Even on the college campus, where academic prerogatives are more clearly established, the primacy of teaching is given little more than lip-service.

Teachers Must Earn Status by Expertness

However, teachers cannot achieve the respect and recognition which they ought to have, and which they *must* have if they are to function as true professionals, if they are content with low standards of performance. They must *earn* status by demonstrating what they can do, by indisputable proof of expertness. We insist that the physician shall be able to accomplish more than the barber and the midwife. Unless the teacher's efforts can produce results distinctively superior to those attainable by the intelligent layman, he has no rightful claim to professional status.

The improvement and further professionalization of teaching will require significant changes in the attitudes of teachers themselves. They must assume much greater responsibility for improving the quality of instruction than they have thus far been willing to accept. Instead of resisting attempts to evaluate the effectiveness of teaching, teachers should be in the vanguard of such efforts, resolving the complex problems of evaluation and establishing more precise means of identifying superior performance.

A major cause of low teacher com-

pensation, and of the patronizing attitudes toward the profession, is the fact that the teacher's contribution is vague and indeterminate. No clear distinction is made between skilled performance and incompetency. Without means of judging the teacher's effectiveness, it is convenient to assume that none are especially effective and to treat them all accordingly. Inadequate teaching on any level, anywhere, weakens the structure of the profession and casts a shadow on the work of all teachers. The good teacher's sole defense is a sound evaluation system whereby excellence can be recognized and shoddiness exposed. Our present scheme of bureaucratic leveling, in both private and public institutions, shields the incompetent and victimizes the superior performer.

No substantial progress will be made toward the elevation of teaching standards without vigorous and sustained efforts within the teaching profession. Legal requirements serve only to set the barest minimums. Administrative officers of educational institutions, even though they are deeply concerned, can have only limited influence upon the quality of instruction. Major responsibility must rest with teachers themselves.

Index

Aberle, D. F., 94, 97
Abraham, H. J., 377, 378
Academic Achievement, Wide Range Test of, 385–389
Academic motivation, measures of, 448–450
Acceleration, effect on pupils, 363–372
Achievement, in elementary school boys, 62–69
  scholastic, discussion of, 447–451
  testing, history of, 20
Achievement, Metropolitan Test of, 431
Adolescents, alienation of, 131–142
Aggression, evaluated by primary grade children, 265–268
Agranoff, B. W., 152, 158
Alexander, E. O., 280, 282
Alienation, of adolescents, 131–142
Allen, K. E., 254, 255, 259, 260
Allison, J. B., 416
Allport, G. W., 11, 32, 71, 76, 161, 163, 181, 453, 458
Alpert, R., 128, 130, 252
Altus, W. D., 89, 97, 427
Anastasi, A., 330, 331, 334
Anastasiow, N., 328–335
Anderson, J. E., 30
Aneshansel, J., 453, 459
Angell, J. R., 13
Anxiety, reduction of, 143–145
Aptitude tests, Flanagan, 366, 367
Aquinas, T., 6
Aristotle, 6
Arithmetic, disability in emotionally disturbed children, 385–389
Army Alpha, 23
Arnsworth, M., 334
Aronfreed, J., 252
Arrowood, C. F., 8
Artley, A. S., 454, 459
Asch, S. E., 493, 494
Associationism, 8
Asymmetrical Preference Test, 347
Atkinson, R. C., 226, 301, 304
Attention-seeking, negative, 246–252
Attitudes, toward school, 270–277
Ausubel, D. P., 196, 202
Authoritarianism, and education, 491–494
Aware, process of becoming, 163–165
Ayres, L. P., 20, 21

Baer, D. M., 252, 254, 259
Bain, A., 9
Bainbridge, J. A., 376, 379
Bakan, R., 409–417
Baker, G. D., 28
Baldwin, B. T., 25
Baldwin, J. M., 11, 15, 16
Bandura, A., 247, 252, 453, 458
Barber, T. X., 464, 478
Barnes, R. H., 416
Barron, F., 346, 347, 350, 493, 494
Bartlett, F. C., 453, 458
Bayer, A. A., 95, 97
Bayley, N., 59, 61, 94, 97
Beatty, W. H., 160–180, 182
Behavior, as viewed by primary grade children, 261–269
  disruptive, 246–252
  modification of, 253–260
  pupil, effect on teacher attitudes toward, 463–479
  teacher ratings of, 209
Behaviorism, 26–27, 32
Behavior modification, 235–245
  in preschool, 253–260
Beller, E. R., 209
Bennett, E. L., 153
Bennett, G. R., 448, 450
Bentzen, F., 64, 69
Bereiter, C., 203, 205, 206, 214, 329
Bereiter-Engelmann Language Training Project, 205, 206

Bernreuter, R. G., 24
Berry, P. C., 345, 351
Berzonsky, M. D., 205, 214
Bexelius, A., 375, 379
Biddle, B., 262
Bieber, I., 96, 97
Bills, R. E., 165, 181
Binet, A., 12, 19
Birch, H., 331, 334, 416
Birth-order, effects of, 88–97
Bish, G. G., 355, 358
Blair, G. M., 15, 32
Blake, R., 453, 455, 458
Blakeley, W. P., 77–87
Blane, W. A., 416
Blank, M., 209
Bledsoe, J. C., 75, 76
Block, C. H., 345, 351
Block size, optimal in learning lists, 223–224
Bloom, B., 59, 61, 294, 300, 304–305
Blos, P., 137, 138, 142
Boring, E. G., 10, 32
Boston, M., 334
Botticelli, S., 346
Bowlby, J., 330, 334
Boykin, L. L., 448, 450
Brain chemistry, 149–159
Brain damage, and malnutrition in children, 414–415
Brainstorming, and creativity, 344–351
Braley, K. W., 453, 459
Brazziel, W. F., 329
Brewer, T. H., 417
Brickell, H. M., 294
Bright, R. L., 125
Brim, O. G., Jr., 94, 97
Briskin, A. S., 253–260
British Infant Schools, 213–214
Brode, E. L., 270–277
Brodsky, G., 254, 259
Brophy, J. E., 463–479
Brown, B. R., 75, 76
Brownell, W. A., 30
Bruner, J. S., 158, 182–193, 196, 202, 303, 304
Buckingham, B. K., 18, 21
Buell, J., 254, 259
Bullock, H. A., 448, 450–451
Burke, B., 22
Burke, C. J., 220, 230
Burton, R. V., 258, 259
Buss, A. H., 117–120
Buswell, G. T., 22, 30
Butterfield, E. C., 215
Butts, R. F., 26, 32
Byrne, D., 34–51

California Personality Inventory, 482, 485
Campbell, D. T., 294
Campbell, J., 345, 350
Carter, C. O., 325
Cattell, J. McK., 9, 19
Cazden, C., 209
Centers, R., 493, 494
Central tendency, measures of, 36–39
Chalupsky, A. B., 493, 494
Chapin, F. S., 294
Cheating, effect of anxiety on, 126–130
Child-centered practices in education, 491–494
Child-rearing practices, and children's achievement, 62–69
Child, C. M., 26
Chomsky, N., 190, 194, 212
Ciavarella, M. A., 373–379
Clark, D., 332, 335
Clark, E. T., 70–76
Clark Doll Test, 454–456
Clark, H. B., 457, 458
Clark, K., 60, 61
Clark, K. B., 426, 427
Clark, M. K., 457, 458
Closed-mindedness and classroom recognition, 278–282
Cobb, S., 96, 97
Coghill, G. E., 26
Cognitive development, and environmental complexity, 149–159
  Karnes curriculum for, 205
  preschool programs in, 205
Coladarci, A. P., 427
Combs, W., 71, 76
Comic books, effect on children's behavior, 77–87
Competence motivation, 188–189
Computer-assisted instruction, 226–230
Computers in contingency management, 242–244
Conant, J. B., 294, 448, 451
Concept formation, in mathematics, 221–223
Contingency contracting, and motivation, 242–244

Contingency contracting (*continued*)
self-management of, 244
teacher resistance to, 240–241
with potential dropouts, 239–240
Contingency management, computerized, 242–244
in preschool, 238–239
Conway, J. A., 278–282
Coombs, A. W., 103, 113
Coopersmith, S., 68, 69
Corot, J. B. C., 346
Correction procedure, overt, 220–221
Correlation coefficient, 45–51
Counselors, school, proper role of, 374–378
Coursin, D. B., 414, 416
Cowdery, K. M., 24–25, 32
Cowley, J. J., 411, 416
Cox, K., 22
Craig, R. C., 196, 202
Crandall, V. C., 68, 69, 252
Cravioto, J., 412–413, 416
Creativity, and brainstorming, 344–351
and intelligence, 355–356
and "orneriness," 346–350
and preferences of art forms, 346–350
in children, 353–358
relation to I.Q., 338
stimulation of, 344–351
teacher assessment of, 353–354, 357–358
teacher response to, 353
tests of, 366, 367
Cremin, L. A., 26, 32
Crick, F., 326
Crissman, P., 13, 32
Cronbach, L., 329, 330
Crothers, E. J., 219, 224
Crow, J. F., 329, 330
Crowne, D. P., 252
Crutchfield, R. S., 158, 346, 350
Cubberly, E. P., 21, 32
Cundrick, B. P., 422–427, 428
Curtis, C. A., 205, 214

Darlington, C. D., 326
Darwin, C., 9, 25
Davidson, H., 75, 76
Davis, K., 247, 252
Davis, D. L., 476, 478
Davis, R. A., 21, 32
Day, J., 8
Dearborn, W. F., 24, 26
DeBaca, P. C., 245
Deboer, J. J., 86
Decarie, T., 58, 61
Defendi, V., 416
De Groat, A. F., 476, 478
De Licardie, E. R., 416
Dellinger, W. S., 416
Democritus, 6
Dempsey, R., 143–145
Dennis, W., 453, 454, 459
Dentler, R., 459
Descartes, R., 55–58
Dettre, J. R., 280, 283
Deutsch, M., 60, 61, 205, 214, 419, 421
Devine, J. A., 245
Dewey, J., 13–14, 24, 30, 31, 230, 333, 335, 337, 341
Dewey, R., 68, 69
Diamond, M., 153
Dickie, J. P., 205, 214
Diener, M. M., 416
Di Lorenzo, L. T., 205, 214
Disadvantaged children, compensatory education for, 329
early education of, 204–214, 428–440
education of, 324
interaction with mother, 418–421
parents of, 204
Discovery method of learning, 182–194
motivating effect of, 195–202
Disturbed children, coping with, 121–125
Dogmatism, Rokeach's scale, 279
Dollard, J., 30
Doolittle, L. W., 373–379
Dorwort, W., 252
Dowley, E. M., 206, 215
Downing, L. N., 427
Draw-a-Man Test, 424, 426, 427
Dressel, P. L., 280, 282
Drive, related to learning, 236
Drive-reduction model of learning, 188
Dropouts, high school, programs to aid, 397–404
Dubin, E., 71, 75, 76
Dubin, R., 71, 75, 76
Dubnick, L., 483
Duncan, D. D., 325
Dunnette, M. D., 345, 346, 348, 350
Dyer, H. S., 298, 304
Dybwad, G., 78, 86

Early experience, 149–159
Ebbinghaus, H., 9, 10
Ebel, R. L., 279, 282, 314–319
Eby, F., 8
Echelberger, E., 104, 113
Eckland, B., 325
Education, improvement of, and educational measurement, 297
Educational psychology, history of, 5–33
Edwards, A. L., 453, 458
Eichenwald, H. F., 413, 416
Electives, benefits of, 337–338
Elementary school curriculum, learning theory in, 217–230
Elkind, D., 329, 331, 334
Elliott, E. C., 20
Emotionally disturbed children, educational disability in, 384–389
Emotional problems, and grades, 390–396
Empiricism, 7
Enburg, R., 385, 389
Englemann, S., 205, 206, 214
Environment, influence on intelligence, 329–333
Erickson, E., 58, 61, 63, 69, 138, 139, 142
Epperson, D. C., 101–113
Erickson, M. T., 413–414, 416
Estes, W. K., 161, 175, 181, 220, 221, 226, 230
Eurich, A. C., 28
Evans, L., 427
Evening school programs, for high school dropouts, 401–402
Ezerman, R., 252

Faculty psychology, 6–8, 13
Farber, N. A., 481
Farquhar, W. W., 447–451
Featherstone, J., 213–214
Fink, M. B., 75, 76
First-born, personality of, 88–97
Fisher, R. A., 326
Fisher, R. L., 114–120
Fiske, M., 79, 87
Fitzgerald, F. S., 410, 415
Fitzgerald, J. A., 427
Fixer, G., 332, 335
Flanagan Aptitude Classification Tests, 366, 367
Flanagan, J. C., 294–295
Flanders, N. A., 270–277
Foley, J. P., 330, 334
Foreign language, learning vocabulary in, 223–224
Frank, J., 79, 86
Franke, P. R., 21, 32
Freeman, F. N., 22
Freeman, H., 295
French, J. R., 96, 97
Freud, A., 137, 142, 330, 334
Freud, S., 55–58
Friedenburg, E. Z., 481, 485
Fromm, E., 161, 181
Fry, P. C., 416
Functionalism, 11, 13
Furer, M., 137, 142

Gage, N. C., 294
Gagne, R. M., 335
Gainsborough, T., 346
Gallimore, R., 246–252, 253, 261, 278
Galton, R. F., 8–9, 11
Gardner, W. I., 253–260
Garth, T., 427
Gates, A. I., 21, 28, 30
Gates, A. J., 416
Gesell, A., 22, 25, 72, 76
Gestalt psychology, 27, 32
Getzels, J. W., 280, 283, 338, 489
Gerard, H. B., 250, 252
Gewirtz, J. L., 252
Gibran, K., 164, 181
Gifted children, effect of acceleration on, 363–372
mathematics education of, 217–219
Gilbert, O., 412, 416
Gilligan, C. F., 127, 130
Ginsberg, R., 222, 230
Glaser, R., 295
Glick, J., 209
Goddard, H. H., 19
Goethe, J. W. von, 337
Gold, M., 104, 113, 476, 478
Goldstein, K., 161, 181
Good, T. L., 463–479
Goodenough, F., 22
Goodenough-Harris Draw-a-Man Test, 424, 426, 427
Goodman, M. E., 453, 458
Goodman, P., 481
Goodwin, W. L., 363–372
Gordon, E. W., 304
Gordon, I. J., 55–61, 71, 76

Gordon, J. E., 416
Grades, relation to emotional problems, 390–396
Grading, influence on college professors, 278–283
Gray, A. W., 63, 65, 69
Gray, J. S., 26, 32
Gray, S. W., 205, 215, 422, 428–440
Gray, W. S., 19, 22, 29, 33
Green, R. L. 447–451
Greenwald, H. J., 457, 458
Gregor, A. J., 453, 454, 457, 458
Greulich, W. W., 330, 334
Group therapy, 380–404
Gruenwald, G., 414, 416
Guba, E. G., 484, 489
Gump, P. V., 261–269
Guntrip, H., 139, 142
Guskin, J., 332, 335
Guthrie, E. R., 30, 175
Guthrie, G. M., 251, 252

Haber, R. M., 128, 130
Habit, 14
Hadassah, D., 453, 459
Hagen, E., 483, 490
Haggerty, L. H., 24
Hall, G. S., 9, 10, 15, 16, 23, 24, 25
Hall, W. B., 346, 350
Halleck, S. L., 141, 142
Hamlin, R. H. J., 416–417
Harrell, R. F., 414, 416
Harrell, W., 26, 33
Harris, F., 254, 255, 259, 260
Harrison, R., 26, 32
Hart, B., 254, 259
Hartmann, G. W., 30
Havighurst, R. J., 72, 76, 427, 481
Hawkridge, D. G., 204, 214
Head Start Programs, 60, 291
Heard, C. R. C., 416
Hebb, D. O., 165, 181
Heinicke, C. M., 330, 334
Heisler, F., 78, 87
Hellyer, S., 220, 230
Hemingway, E., 410, 415
Hemphill, J. K., 295
Hemmon, V. A. C., 24
Henderson, E., 61
Henri, V., 12
Herbart, J. F., 8, 21, 24
Herrick, C. J., 26
Hertzig, M. E., 331, 334
Herzog, E., 295
Hess, R. D., 331, 334, 419, 421
Hesseling, P., 295
Heuristic approach to learning, 182–194
Hewett, F. M., 254, 259
Hildreth, G., 22
Hilkevitch, R. R., 427
Hill, J. P., 126–130
Hobbs, N., 63, 69
Hodges, W. L., 205, 214
Hodgins, A. S., 205, 214
Hoehn, A. J., 476, 478
Horn, E., 30
Hollingworth, L., 22
Holt, J., 481
Holzinger, K. J., 22
Homme, L. E., 235–245, 253
Hood, L. E., 486, 489
Hopkins, T. K., 295
Horowitz, E., 453, 458
Horowitz, L., 454, 458
Horowitz, R. E., 453, 454, 458
Hostility, maternal, 117–120
Hoult, T. F., 79, 87
Howell, R. J., 427
Hull, C. L., 30, 170, 175, 181
Hunt, J. McV., 61, 329, 330, 332, 333, 334, 335
Hyman, H. H., 295

Identity, development of, 162–163
Ilg, F., 72, 76
Independence, and education, 491–494
in elementary school boys, 62–69
Indians, American, measuring intelligence of, 422–427
Individual differences, 337–343, in learning, 217
Infants, interaction with mother, 418–421
Infant Schools, British, 213–214
Inference, statistics of, 41–45
Inhelder, B., 330–331, 333–334, 335
Innate ideas, theory of, 7
Innovation, in secondary schools, 143–145
In-service training for teachers, in designing tests, 318–319
Instrumentalism, 13
Intelligence, and American Indian children, 422–427

Intelligence (*continued*)
and creativity, 355–356
and malnutrition, 411–413
and racial differences, 325–327, 328–335
and social class differences, 325
effects of childhood deprivation on, 330–333
influence of environment on, 329–333
testing, history of, 11–13, 23
testing of preschool children, 207–208
tests, Wechsler Scale for Children, 385–386, 423–425, 427
Wechsler Pre-School Test of, 423, 424, 427
Interpersonal relations, and mental health, 101–113
IQ, computation of, 41
Kuhlman-Anderson Test of, 365
Isaacs, S., 30
Ivanov-Smolensky, A. G., 190, 193

Jaastad, K., 345, 350
Jackson, P. W., 280, 283, 338, 376, 379, 476, 478, 489
Jacobs, J. M., 448, 451
Jacobs, P. J., 251, 252
Jacobson, L., 427, 464, 479
Jahoda, M., 102, 113
James, W., 10–11, 17, 56
Jastak, J., 387–388, 389
Jencks, C., 204, 214
Jensen, A. R., 323–327, 329–334
Jersild, A. T.., 30, 32, 71, 76
Johnson, B. C., 453, 459
Johnson, D. W., 452–459
Jones, E. E., 250, 252, 453, 459
Jones, H. E., 22, 94, 97
Johnston, M., 254, 259, 260
Joseph, A., 427
Josselyn, I. M., 247, 252
Judd, C. H., 18–19

Kagan, J., 59, 61, 329, 332, 333, 335, 418–421, 428
Karnes, M. B., 205, 214
Katkovsky, W., 68, 69, 252
Katz, D., 453, 459
Katz, M., 302–303, 304
Kaufman, B., 247–248, 252
Keith, R. A., 63, 69
Kelley, C., 254, 259
Kelley, T., 22
Kemp, B., 248, 252
Kennedy, W. A., 414, 416
Kersh, B. Y., 195–202
Kessler, J., 63, 66, 69
Kirkpatrick, E. A., 25
Kittell, J. E., 196, 202
Klaus, R. A., 205, 215, 422, 428–440
Klausmeier, H. J., 363–372
Klein, R. E., 412, 416
Koch, H., 95–97
Kochendorfer, R. A., 127, 130
Köhler, W., 27
Koffka, K., 27
Kohl, H., 481
Kohlberg, L., 209, 212, 215
Kotz, N., 417
Kounin, J. S., 261–269
Kozol, J., 481
Krasnogorsky, N. D., 190, 193
Krathwohl, D. R., 300, 304–305
Krech, D., 149–159
Krugman, M., 14, 33
Kuethe, J., 115–120
Kugelmass, N., 414, 416
Kuhlmann-Anderson IQ Test, 365
Kuhlmann, F., 19
Kushlick, A., 416

Lahaderne, H. M., 476, 478
Lambert, W. E., 225, 230
Landreth, C., 453, 459
Lang, G., 75, 76
La Rochefoucauld, F. de, 337
Learning, by directed discovery, 195–202
capacities of lower-class children, 332–333
difficulties, in emotionally disturbed children, 384–389
drive-reduction model of, 188
evaluation of, systems approach, 292
evaluative research in, 287–295
instrumental, 172–176
intrinsic, 172–176
motivational effect of, 201–202
rote method, 198, 200
self-concept theory of, 161–181
studies of, 29–30
theory, related to elementary school curriculum, 217–230
Lehmann, H. C., 30
Lehmann, I. J., 280, 283
Leiter International Performance Scale, 208

Lesser, G., 332, 335
Levin, H., 249, 252, 331, 334
Levine, J. M., 453, 459
Levine, L. S., 480–490
Levy, D. M., 64, 69
Lewes, G. H., 9
Lewis, K., 27, 30
Lewis, M., 252
Li, J. C. R., 198–199, 202
Lindgren, F., 344–351
Lindgren, H. C., 34–51, 344–351, 491–494
Lindquist, E. F., 431–432, 440
Lipsitt, L., 56, 61
Lippitt, R., 30, 104, 113, 476, 478
Litcher, J. H., 452–459
Liversant, S., 274, 277
Locke, J., 7
Loevinger, J., 299, 301, 305, 331, 334
Long, B., 61
Longenecker, E. E., 165, 181
Lorge, I., 28, 33
Lott, A. J., 448, 451
Lott, B. E., 448, 451
Lowe, C. U., 415
LSD, use by adolescents, 132–133, 136
Ludeman, W. W., 427
Luszki, M. B., 101–113
Luria, A. L., 190, 193

McAndrew, G. L., 374, 379
McCall, W., 22
McCandless, B. R., 71, 76, 205, 214
McConnell, J., 152–153, 159
McConnell, T. R., 30
McCuen, J. T., 64, 69
McGeach, J. A., 16, 33
McGough, J., 151, 153, 159
McGuire, C., 165, 181
McLellan, J. A., 230
McNeill, D., 158, 159, 332, 334
McPherson, D. A., 453, 454, 457, 458
MacDonald, S., 248, 251, 252
Maccoby, E. E., 249, 252, 331, 334
Macfarlane, J. W., 89, 97

Mahler, M. S., 137, 142
Maimonides, 183, 193
Maller, J. B., 30
Malnutrition, and intelligence, 411–413
  effect on cognitive functioning, 409–417
Malpass, L. F., 113
Manning, W. H., 296–305
Marans, A., 60, 61
Markey, F. V., 31
Marlowe, D., 252
Masia, B. B., 304–305
Maslow, A. H., 161, 181
Mathematics, in elementary school curriculum, 217–219, 225–228
Mather, K., 326
Mead, G. H., 13, 61
Mead, M., 415, 417
Meade, J. E., 416
Mean, 38–39
Measurement, educational, uses of, 298–299, 314–319
Median, 38–39
Mediation, generic, 192
  part-whole, 193
  thematic, 192–193
Mees, H., 254, 260
Memory, Ovsiankina effect, 202
  Zeigarnik effect, 202
Mendez, O. A., 331, 334
Mental health, in the classroom, 101–113
  of boys, 114–120
Mentally retarded, educable, 380–383
Merton, R. K., 252
Messick, S., 305
Metcalf, D., 416
Meumann, E., 16
Meyer, W. J., 476, 478–479
Michigan State M Scales, 448–450
Miles, M. B., 295
Miller, G. A., 192, 194
Miller, J. O., 431, 440
Miner, J. B., 24
Mitzel, H. E., 483
Modigliani, A., 346
Monckeberg, F., 411, 416
Monroe, M., 454, 459
Monroe, P. A., 21, 33
Monroe, W. S., 21, 22, 29, 33
Montaigne, M. de, 304
Montessori, M., 341
Moore, A. U., 416

Moore, B., 24
Moore, K. C., 25
Moore, W. E., 295
Morland, J. K., 459
Morris, N. L., 11, 33
Morrison, B. M., 270–277
Morrison, D. C., 255, 259–260

Moss, H., 59, 61
Mother-child interaction, 418–421
Motivation, academic, 447–451
  and contingency contracting, 242
  differential probability hypothesis, 237–238
  related to learning, 236–237
Motivational states, relation to self-concept, 167–168
Mowrer, O. H., 163, 164, 175, 181
Muecher, H., 414, 416
Munn, N. L., 29, 33
Munro, H. N., 416
Murphy, G., 58, 453, 459
Murphy, L., 30, 59, 61

Naegle, K. D., 94, 97
Naeye, R. L., 416
Negative attention seeking, 246–252
Negroes, children's attitudes toward, 452–459
Nelson, R. H., 487, 489–490
Newton, I., 183, 251
Norton, D. P., 448, 451
Noyes, K. J., 374, 379
Nutrition, prenatal, and intelligence of offspring, 413–414

Ogden, R. M., 27, 33
Olds, J., 172–173, 179, 181
Ombudsman, model for school counselor, 374–378
Open-mindedness, and classroom recognition in college, 278–283
Oppenheim, D. B., 457, 458
Orzeck, A. Z., 165, 181
Osborn, A. F., 345, 351
Otis, A., 23
Ovsiankina effect, 202
Ozehosky, R. J., 70–76

Palmer, O., 279, 282
Parental expectations, for boys, 62–69
Park, D. G., 27, 33
Parke, R. D., 250, 251, 252
Parkes, A. S., 416
Patterson, G. R., 254, 259
Patton, G. M., 493, 494
Pavenstedt, E., 60, 61, 419, 421
Pavlov, I., 189–190
Payne, D. A., 451
Peabody Picture Vocabulary Test, 208, 424, 426, 427, 431, 432
Pearson, K., 9
Penfield, D. J., 427
Perception, Ames demonstrations of, 161
Perkins, C., 76
Permissiveness, in education, 491–494
Personality tests, Edwards Personal Preference Test, 482
  Minnesota Multiphasic Inventory, 482
Pestalozzi, J. H., 7
Peterson, J., 22, 24
Peterson, R. F., 255, 259–260
Piaget, J., 57, 58, 61, 158, 203, 300, 330–331, 333–334, 335
Picasso, P., 346
Pintner, R., 24
Piskula, G., 366
Plato, 6
Platt, B. S., 416
Platt, J. R., 336–343
Pond, W. G., 416
Postman, L. J., 453, 458
Potter, R. G., Jr., 97
Poull, L. E., 416
Pragmatism, 13
Praise, as behavior modifying agent, 256, 257–259
  related to pupil attitudes, 270–277
Prejudice, racial, in children, 452–459
Premack, D., 237–239, 245
Prenatal nutrition, and intelligence of offspring, 413–414
Preschool, cognitive oriented curriculum, 206–208
  comparison of curricula, 203–215
  for disadvantaged children, 428–440
  importance of language development in, 212
  involvement of mother in, 211, 213
  language training curriculum, 206–208
  staff model for, 210–213
  unit-based curriculum, 206–208
  Ypsilanti Demonstration Project, 206–214
Pressey, J. L., 24
Pressures, reduction of, 143–145
Preston, A., 68, 69
Probability, 44
Problems of adjustment, of Latin-American students, 442–446
Problem-solving, education for, 303
  process of, 191–192
Progressive education, 28–29

Projective tests, of social distance, 114–120
Provocation, tactical, form of misbehavior, 248–252
Provus, M., 295
Psychoanalysis, 27, 32
Psycholinguistic abilities, Illinois Test of, 205
Pumpian-Mindlin, E., 138, 142
Punishment, related to learning, 236
Pupil attitudes, and mental health, 102–103
  toward school, 270–277
Pupil Behavior Inventory, 209
Pygmalion in the Classroom, discussion of, 464

Questionnaires, invention of, 10

Race, children's attitudes about, 452–459
  tests of children's attitudes about, 454–458
Radke, M. J., 453, 459
Rau, L., 252
Read, I. M., 416
Readers, multiethnic, use of, 452–459
Readiness, testing for, 72–73
Reading, disability, in emotionally disturbed children, 385–389
  initial phases, 219
Redl, F., 262, 269
Reinforcement, immediate, with primary grade children, 220–221
Reiss, A. J., Jr., 273–277
Response latency, 225–226
Retarded children, work experience programs for, 380–383
Ribonucleic acid (RNA), 150, 153
Rice, J. M., 16, 20, 33
Rickert, E. J., 245
Riesman, D., 191
Ripple, R. E., 364, 372
Risley, T., 209, 254, 259, 260
Roberts, A. O., 204, 214
Roberts, J. M., 97
Robinson, B. W., 156, 157, 159
Robinson, H. M., 454, 459
Robles, B., 416
Rogers, C. R., 71, 76, 161, 165, 181
Rokeach, M., 279, 280, 281, 282, 283
Ronda, T., 363–372
Rosenberg, B. G., 97
Rosenbluth, D., 334
Rosenhan, D., 252
Rosenthal Effect, 212
Rosenthal, R., 426, 427, 464, 479
Rosenzweig, M. R., 153
Rosso, P., 416
Rotter, J. B., 274, 277
Rowland, H., 376, 379
Rowley, V. N., 384–389, 390
Rubinfine, D. L., 137, 142
Rulon, P. J., 28
Ryans, D., 30
Ryans, D. G., 488, 490

Sagi, P. G., 97
Salter, R., 205, 214
Sampling, 41–44
Samuel, E. L., 416
Sandiford, P., 30
Sassenath, J. M., 202
Satisfaction, definitions of, 168–169
Schachter, S., 89, 97
Schaeffer, E. S., 331, 334
Schiller, J. C. F., 337
Schlosberg, H., 224, 230
Schmuck, R. A., 101–113
Scott, P. M., 258, 259
Scrimshaw, N. S., 416
Seashore, H. G., 448, 450
Sears, P. S., 206, 215
Sears, R. R., 57, 61, 165, 170, 181, 249, 252, 331, 334
Sechrest, L., 249–251, 252
Seeman, M., 274, 277
Seifert, K., 215
Self, beginnings of, 55–61
Self-actualization, 161
Self-concept, 55–61, 70–76
  learning theory of, 161–181
Self-esteem, of students, 107–109
Self-expression, model of, 171–172
Self-realization, 161
Seltsam, K., 27
Shannon, B., 76
Sharp, S. E., 12, 33
Shaw, M. C., 64, 69
Sheldon, E. B., 295
Shelton, J., 126–130
Sherwood, C. C., 295
Sheviakov, G., 262, 269
Shinn, M., 25
Shipman, V. C., 331, 334, 419, 421
Siegel, S., 93, 97
Significance level, 44–45

Silver, M. J., 464, 478
Simon, T., 12, 19
Singer, D., 453, 459
Skeels, H. M., 330, 334
Skinner, B. F., 175
Skodak, M., 330, 334
Slobodian, J. J., 476, 478
Slotkin, H., 397–404
Smelser, W. T., 88–97
Smilansky, S., 58, 60, 61
Smith, O., 427
Smythe, M., 412, 416
Snider, J. G., 427
Snow, R. E., 464, 479
Snow, R. F., 212
Snow, R. H., 495–502
Social class, and infants' interaction with mother, 418–421
Social psychology, 30–31
Spanish-speaking students, problems of, 441–446
Speech, relation to early learning, 189–190
Spence, K. W., 175, 190, 193–194
Spencer, H., 9
Sperzel, E. Z., 78, 87
Spicker, H. H., 205, 214
Spitz, R., 61, 137, 142, 330, 334
Standard deviation, 42–44
Stanford-Binet Intelligence Test, 207–208
Stanley, J. C., 294
Starch, D., 20
Statistics, introduction to, 34–51
  of description, 35–36
Steffens, L., 191, 194
Steinhorst, R., 254
Stern, W., 19
Stewart, L. H., 88–97
Stewart, R. J. C., 416
Stimulus generalization, 221
Stoch, M. B., 412, 416
Stone, F. B., 384–389, 390
Strong, E. R., 24–25
Stroud, J., 29, 30, 388, 389
Student assessment, in college classes, 278–283
Students, Negro, 448–451
Suchman, E. A., 287–295
Sullivan, H. S., 56, 61
Sully, J., 8
Suppes, P., 216–230
Sutton-Smith, B., 96, 97
Symonds, P. M., 202

Taba, H., 333, 335
Tamkin, A. S., 385–389
Taylor, C., 103, 113
Taylor, D. W., 345, 351
Teacher expectations of classroom performance, 463–479
Teacher ratings of preschool children, 209
Teachers, non-punitive, effect on pupil behavior, 262–269
  personality characteristics of, 480–490
  punitive, effect on pupil behavior, 262–269
  training of, 480–490
Teacher-student relations, 176–179
Teaching, anxieties in, 495–502
  motivation for entering, 486–489
  psychological view of, 480–490
  self-evaluation of, 497–498
Tenenbaum, S., 390–396
Terman, L. M., 19–20, 22, 25
Teska, J. A., 205, 214
Tests, achievement, 310
  correlation between, 312–313
  derived scores, 311
  intelligence, 308–309
  in vocational choice, 309
  educational, 306–313, as instruments of educational theory, 299–301, design of, 315–318, external validity, 299, guidance uses of, 301–303, misuse of, 306–313, structural validity, 299, substantive validity, 299, traditional uses of, 298–299
Tharp, R. G., 248, 251, 252
Therapy, group, 390–396
Thomas, A., 331, 334
Thompson, G. G., 476, 478
Thompson, M. L., 487, 489–490
Thorndike, E. L., 15, 17–18, 20, 22, 24, 25, 32, 33
Thorndike, R. L., 464, 479, 483, 490
Thresher, B. A., 302
Thurstone, L. L., 23, 24
Thurstone, T. G., 24
Tiedeman, D., 303
Tift, K. F., 121–125
Time-out, as behavior modification method, 254–260
Titchener, E. B., 12
Tolman, E. C., 156, 159

Torrance, E. P., 355, 358
Toulouse-Lautrec, H. de, 346
Trager, H. G., 453, 459
Transfer of learning, 221–223
Travers, R., 28
Trow, W. C., 30
Tuddenham, R. D., 493, 494
Tulkin, S. R., 418–421, 428
Turner, G. H., 427
Turner, M., 380–383, 397
Tyler, F. B., 113
Tyler, R. W., 28, 294, 295, 333, 335

Van Egmond, E. E., 104, 113
Variability, measures of, 40–41
Vigotsky, L. S., 189, 193
Vocational counselling, for high school dropouts, 398–404

Wahler, R. G., 255, 259
Wall, A. M., 252
Wallace, J., 249–251, 252
Walters, H., 453, 459
Walters, R. H., 247, 250, 251, 252
Warshauer, M. E., 459
Watson, G. B., 26
Watson, R. I., 5–33
Wattenberg, W., 73, 76
Wechsler Intelligence Scale for Children, 385–386, 423–425, 427
Wechsler Pre-School Intelligence Test, 423–424, 427
Weikart, D. P., 203–215
Weinstern, L., 117, 120
Weir, R., 224
Wertham, F., 78, 87
Wertheimer, M., 27
Wesman, A. G., 448, 450
Westoff, C. F., 95, 97
Wetzel, R., 248, 251, 252
White, R., 188, 193
White, R. K., 30
White, R. W., 63, 69, 161, 181
Whiting, A., 62–69
Willey, R. L., 487, 490
Wilkerson, D. A., 304
Williams, F. E., 352–358
Williams, F. S., 139, 142
Willoughby, R. R., 24
Wilson, L., 305
Wilson, L. N., 25
Winick, M., 411, 416
Winkel, G. H., 255, 259–260
Winterbottom, M. R., 65
Wise, L. J., 131–142
Wissler, C., 12, 33
Witherspoon, R. P., 441–446
Witty, P. A., 30, 78, 87
Wolf, M., 254, 259, 260
Wolf, N., 254, 259
Wolfe, K. M., 79, 87
Wollin, J. E. W., 19
Women, in academic life, 338–339
Womer, F. B., 306–313, 315
Wood, B., 25
Woodrow, H. H., 24
Woods, E., 28
Woodworth, R. S., 17, 224, 230
Woodyard, E. R., 416
Work experience for educable mentally retarded, 380–383
Wright, C. R., 295
Wrightstone, J. W., 28
Wundt, W., 9, 11, 18
Wylie, R., 71, 76, 165, 181

Yamamoto, K., 355, 358
Yarrow, L., 60, 61
Yarrow, M. R., 258, 259
Yoakum, C., 24
Ypsilanti Preschool Curriculum Demonstration Project, 206–214

Zajonc, R. B., 187, 193
Zeigarnik Effect, 202
Zigler, E., 212, 215