

Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure?

TRACI SITZMANN
University of Colorado, Denver

KATHERINE ELY
Fors Marsh Group

KENNETH G. BROWN
University of Iowa

KRISTINA N. BAUER
Old Dominion University

We conducted a meta-analysis to clarify the construct validity of self-assessments of knowledge in education and workplace training. Self-assessment's strongest correlations were with motivation and satisfaction, two affective evaluation outcomes. The relationship between self-assessment and cognitive learning was moderate. Even under conditions that optimized the self-assessment–cognitive learning relationship (e.g., when learners practiced self-assessing and received feedback on their self-assessments), the relationship was still weaker than the self-assessment–motivation relationship. We also examined how researchers interpreted self-assessed knowledge, and discovered that nearly a third of evaluation studies interpreted self-assessed knowledge data as evidence of cognitive learning. Based on these findings, we offer recommendations for evaluation practice that involve a more limited role for self-assessment.

To remain competitive in today's dynamic environment, organizations must have employees who maintain up-to-date knowledge and skills. Workplace training and educational programs both play an important role in this effort. A challenge in these areas of practice is that it can be time consuming and expensive to develop metrics to assess knowledge levels. One quick and efficient means of assessing knowledge is the use of self-report measures. *Self-assessments* of knowledge are learners' estimates of how much they know or have learned about a particular domain. Self-assessments offer the potential to reduce the burden of developing

tests to determine whether the desired knowledge has been gained as a result of participation in a course or training intervention.

Prior research, as well as age-old wisdom, suggests that self-assessments of knowledge may have limitations. In 1750, Benjamin Franklin proposed, "There are three things extremely hard: steel, a diamond, and to know one's self." In addition, Charles Darwin (1871) noted, "Ignorance more frequently begets confidence than does knowledge." These quotes are consistent with Kruger and Dunning's (1999) research findings that some people routinely overestimate their capabilities. Similarly, accrediting bodies, such as the Association to Advance Collegiate Schools of Business (AACSB) International, require schools to provide evidence of student learning as part of the accreditation process and recommend directly assessing learning rather than relying on student self-assessments (AACSB International, 2007).

This work was funded by the Advanced Distributed Learning Initiative, Office of the Deputy Under Secretary of Defense for Readiness and Training, Policy and Programs. The views expressed here are those of the authors and do not necessarily reflect the views or policies of the Department of Defense. We would like to thank the editor, Ben Arbaugh, and three anonymous reviewers for their valuable feedback.

Despite the potential limitations of self-assessments of knowledge, they are used as an evaluation criterion across a wide range of disciplines, including education, business, communication, psychology, medical education, and foreign language acquisition (e.g., Dobransky & Frymier, 2004; I-TECH, 2006; Lim & Morris, 2006). Moreover, self-assessments of knowledge are included in many higher education end-of-semester course evaluations to examine teacher effectiveness (Marsh, 1980, 1983).

Our purpose was to compare and contrast research evidence on the validity of self-assessments of knowledge with how self-assessments are used and interpreted in research. Specifically, we examined three research questions. First, how closely are self-assessments related to affective and cognitive learning outcomes? Our categorization of learning outcomes is guided by Kraiger, Ford, and Salas (1993), who classified knowledge test scores as cognitive learning outcomes and learner motivation and self-efficacy as affective learning outcomes. Reactions are also an affective outcome and reflect satisfaction with the learning experience. Extensive research has examined the relationship between self-assessment and cognitive learning (e.g., Dunning, Heath, & Suls, 2004; Falchikov & Boud, 1989), but research has not systematically examined the degree to which learners' self-assessments correlate with reactions, motivation, and self-efficacy. Understanding the construct validity of self-assessments relative to other commonly used evaluation outcomes will clarify the role that self-assessments of knowledge should have in educational and workplace training evaluation efforts. To address these questions, we reviewed the literature and conducted a meta-analysis of 222 independent samples, including data from more than 40,000 learners.

Second, do course design or methodological factors influence the extent to which self-assessments correlate with cognitive learning? It is possible that self-assessments may have large relationships with cognitive learning under certain learning conditions. For example, programs that offer extensive feedback should provide learners with the information necessary to accurately calibrate their self-assessments (Butler & Winne, 1995; Kanfer & Kanfer, 1991). The current study examined several potential moderators of the self-assessment-learning relationship: (1) whether learners received externally generated feedback on their performance; (2) whether the delivery medium was classroom instruction, blended learning, or Web-based instruction; (3) whether, based on the nature of the course content, the

course may have provided natural feedback by way of the consequences of learners' actions; (4) whether learners practiced rating their knowledge levels and received feedback on the accuracy of their self-assessments of knowledge; and (5) the nature of the self-assessment questions. These moderator analyses may provide insight into ways to strengthen the relationship between self-assessments and knowledge.

Third, how are self-assessments of knowledge used and interpreted in evaluation research? We reviewed published and unpublished evaluation studies to determine whether empirical evidence on the validity of self-assessments has influenced the use and interpretation of self-assessment data over time. Specifically, as research has been published indicating that learners may be inaccurate in rating their own knowledge (e.g., Witt & Wheelless, 2001), have researchers changed how they interpret self-assessment measures? In addition, we examined whether the interpretation of self-assessments of knowledge differs across disciplines. Specifically, relative to other disciplines, how are researchers in the business literature interpreting self-assessments?

Prior meta-analytic research provides only partial answers to these questions. Falchikov and Boud (1989) examined the correlation between student and faculty ratings of performance on a variety of course-related outcomes, ranging from course grades to grades on particular assignments. Their meta-analysis provides some insight into the magnitude of the relationship between self-assessments and student performance, but it is now dated and did not employ modern meta-analytic practices, most notably corrections for statistical artifacts. Three previous meta-analyses that are consistent with our broad focus on education and organizational training are Mabe and West (1982), Alliger, Tannenbaum, Bennett, Traver, and Shotland, (1997), and Sitzmann, Brown, Casper, Ely, and Zimmerman (2008). Mabe and West (1982), however, included children in their sample and studies where learners self-assessed their ability (i.e., general intelligence and scholastic ability). We limited our focus to adult populations (over age 18) and only included studies where learners were self-assessing their knowledge of the material taught in a course or training program. The other two meta-analyses, Alliger et al. (1997) and Sitzmann et al. (2008), examined relationships among a variety of course outcomes, including reactions and learning. However, neither study included self-assessments of knowledge as a construct in the meta-analysis.

Another potential area of overlap is with re-

search on student evaluations of teaching (SET). Extensive research has been conducted to examine whether SET are related to cognitive learning (e.g., Ambady & Rosenthal, 1993; Arreola, 2006; Centra & Gaubatz, 1998; Cohen, 1981; d'Apollonia & Abrami, 1997; Feldman, 2007; Greenwald, 1997; McKeachie, 1997). However, SET research examines the relationship between perceptions of teaching behaviors and activities with cognitive learning. Our study does not specifically address this issue because we examine the relationship between self-perceptions of knowledge and cognitive learning. Thus, we are not concerned with teaching and pedagogy directly but with the relationships among outcomes used to assess their effects.

Definitions of Constructs

Self-assessments of knowledge refer to the evaluations learners make about their current knowledge levels or increases in their knowledge levels in a particular domain. Similar to self-assessing job performance (e.g., Campbell & Lee, 1988), when learners evaluate their knowledge, they begin with a cognitive representation of the domain and then judge their current knowledge levels against their representation of that domain. Self-assessments of knowledge are typically measured at the end of a course or program by asking learners to rate their perceived levels of comprehension (e.g., Walczyk & Hall, 1989), competence (e.g., Carrell & Willmington, 1996), performance (e.g., Quiñones, 1995), or increase in these constructs (e.g., Arbaugh, 2005). For example, Zhao, Seibert, and Hills (2005) asked students to rate how much they had learned about entrepreneurship during their MBA education.

We distinguish between self-assessments of knowledge, cognitive learning, and affective outcomes. *Cognitive learning* refers to the understanding of task-relevant verbal information and includes both factual and skill-based knowledge (Kraiger et al., 1993). The critical distinction between self-assessments and cognitive learning is the source that provides the rating of learners' understanding. Self-assessments refer to learners' self-reports of their knowledge levels, whereas cognitive learning refers to grades on exams and assignments as well as instructor ratings of performance.

Three affective outcomes (Brown, 2005; Kraiger et al., 1993)—reactions, motivation, and self-efficacy—are also examined. *Reactions* reflect learners' satisfaction with their instructional experience (Sitzmann et al., 2008). *Motivation* refers to the degree to which learners strove to apply the knowledge

they gained (Sitzmann et al., 2008), whereas *self-efficacy* is learners' confidence in their ability to perform training-related tasks (Bandura, 1977). In the following sections, we discuss previous research on the relationships between self-assessments of knowledge and both cognitive and affective outcomes.

Relationship of Self-Assessments with Cognitive Learning and Affective Outcomes

To understand the validity of a measure, Cronbach and Meehl (1955) suggested researchers must specify the network of constructs related to it and empirically verify the relationships. Empirical evidence suggests that the relationship between self-assessed knowledge and cognitive learning is small to moderate (for definitions of effect sizes see Cohen, 1988). Falchikov and Boud (1989) compared the grading of instructors with undergraduate and graduate students' self-assessments and reported a meta-analytic uncorrected correlation of .39 ($k = 45$, $N = 5,332$). Chesebro and McCroskey (2000) examined the relationship between a 2-item self-assessment measure and a 7-item measure of factual knowledge; they reported a correlation of .50. However, Witt and Wheelless (2001) found the same self-assessment measure only correlated .21 with a longer and more reliable 31-item test of factual knowledge.

Basic psychological research suggests that the correlation between self-assessed knowledge and test performance may be low because some learners are inaccurate. Across multiple domains, Kruger and Dunning (1999) demonstrated that less competent individuals inflated their self-assessments more than highly competent individuals. Participants performing in the bottom quartile on measures of humor, logical reasoning, and English grammar consistently rated themselves above average. These less competent individuals also failed to accurately assess the competence of others. Thus, although the relationship between self-assessments and actual performance was generally positive, self-assessments were a weak and imperfect indicator of actual performance.

In contrast, research suggests that the relationship between self-assessed knowledge and affective outcomes is moderate to large. Because self-assessments are judgments that learners must render, theory suggests that these judgments are influenced by learners' affective and motivational states (e.g., Weiss & Cropanzano, 1996). Baird (1987) examined the extent to which reactions were related to perceptions of learning (across 50 courses with data from 1,600 learners). He found self-

assessments of knowledge were strongly correlated with overall course satisfaction and ratings of the course instructor. Moreover, Carswell (2001) evaluated 11 on-line, work-related courses and found self-assessments of knowledge correlated .55 with motivation. Krawitz (2004) trained 415 clinicians to work with people with borderline personality disorder. He found posttraining self-efficacy correlated .57 with self-assessments of theoretical knowledge and clinical skills. Thus, consistent with theory, self-assessments of knowledge have been shown in prior research to have moderate to large correlations with affective outcomes. Moreover, these correlations are generally stronger than the relationship between self-assessments and cognitive learning.

Hypothesis 1: Self-assessments of knowledge will be more strongly related to affective than cognitive learning outcomes.

Moderators of the Self-Assessment-Cognitive Learning Relationship

It is possible that the strength of the relationship between knowledge and self-assessed knowledge varies based on the assessment context. Therefore, we examined whether various forms of feedback, the delivery media, and the nature of the self-assessment questions moderated the self-assessment-cognitive learning relationship.

External Feedback

Feedback provides learners with descriptive and evaluative information about their performance (Bell & Kozlowski, 2002), and self-regulation research has identified feedback as a critical factor in fostering stronger self-assessment-cognitive learning relationships (Butler & Winne, 1995; Kanfer & Kanfer, 1991; Ley & Young, 2001). When self-assessing their knowledge, learners gather information from multiple sources, including explicit performance feedback provided by others (Kanfer & Kanfer, 1991). If learners receive feedback on their performance, they should modify their self-assessments to be more aligned with their actual knowledge levels, thereby increasing the relationship between self-assessments and cognitive learning.

Hypothesis 2: The relationship between self-assessments of knowledge and cognitive learning will be stronger in courses that provide external feedback on learners' performance during the course than in courses that do not provide external feedback.

Delivery Media

In contrast to Web-based instruction, face-to-face classroom instruction and blended learning (i.e., a combination of Web-based and classroom instruction) generally provide learners with more opportunities to observe the behavior of their instructor and other students. Social comparison theory suggests that individuals strive to maintain accurate views of themselves and use other individuals' points of reference in the absence of an objective standard (Festinger, 1954). When learners detect a discrepancy between their self-assessments of knowledge and desired performance, it triggers search behaviors and may result in learners focusing their attention on the behavior of others (Bandura, 1977). In classroom instruction and in the classroom component of blended learning courses, learners have increased opportunities to observe others. These observations should help them to more accurately calibrate perceptions of their current knowledge levels, increasing the relationship between self-assessment and cognitive learning.

Hypothesis 3: The relationship between self-assessments of knowledge and cognitive learning will be stronger in classroom instruction and blended learning than in Web-based instruction.

Course Content

Educational skills and tasks can be classified into three broad categories: cognitive, interpersonal, and psychomotor (Arthur, Bennett, Edens, & Bell, 2003; Fleishman & Quaintance, 1984). Cognitive skills include understanding the course content, generating ideas, and problem solving. Interpersonal skills involve interacting with others, including public speaking and giving performance feedback to one's subordinates. Psychomotor skills involve performing behavioral activities related to the job, such as typing, repairing the brakes on an automobile, and driving a truck.

These skills may differ in terms of the extent to which learners receive task-generated feedback while learning the course material. Interpersonal and psychomotor tasks provide more task-generated feedback than cognitive tasks. When interacting with other learners and their instructor in interpersonal skills courses, learners receive natural feedback on their performance as they observe others' verbal and nonverbal reactions to it. Psychomotor tasks also provide natural feedback given that learners can observe whether their actions were successful (e.g., do the brakes work?). In

contrast, cognitive tasks do not necessarily afford natural feedback by way of consequences, providing less information to assist learners in aligning their self-assessments of knowledge with their cognitive learning levels. Thus, in courses that teach interpersonal and psychomotor tasks, the self-assessment–cognitive learning relationship should be stronger than in courses that focus on cognitive tasks.

Hypothesis 4: The relationship between self-assessments of knowledge and cognitive learning will be stronger in courses targeting interpersonal and psychomotor skills than in courses targeting cognitive skills.

Practice Self-Assessing Their Knowledge and Feedback on Accuracy

One approach for helping learners calibrate their self-assessments is to provide them with multiple opportunities to self-assess their knowledge and feedback on the accuracy of their self-assessments. For example, Radhakrishnan, Arrow, and Snizek (1996) had learners self-assess their knowledge before and after taking three in-class exams. Their results suggested learners were more accurate in rating their self-assessments of knowledge for the second and third exams than for the first exam. These results are consistent with Levine, Flory, and Ash's (1977) conclusion that self-assessment may be a skill that improves with experience and feedback.

Hypothesis 5: The relationship between self-assessments of knowledge and cognitive learning will be stronger in courses where learners make multiple self-assessments and receive feedback on their self-assessment accuracy.

Similarity of Measures

We also examined the degree of similarity between measures of self-assessment and cognitive learning. Similarity is defined as the degree of congruence, in terms of the level of specificity (e.g., overall knowledge versus specific indicators of declarative and procedural knowledge) and similarity of the constructs assessed (e.g., declarative or procedural knowledge), between the self-assessment and cognitive learning measures. For example, Quiñones (1995) used similar measures by examining the relationship between self-assessments of performance on an exam and objective scores on the same multiple-choice exam.

Conversely, Abramowitz (1999) used dissimilar measures when examining the relationship between learners' self-assessments of their mathematics ability (a general measure of mathematics skills) and their scores on a final statistics exam (a specific measure of statistics knowledge and performance).

Within the personality and performance appraisal domains, researchers have emphasized the benefits of matching the specificity of predictors and criteria (e.g., Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Stewart, 1999) and of providing similar scales for supervisor and self-assessment ratings (Campbell & Lee, 1988). Relationships are strongest when predictor and criterion measures are matched in terms of specificity and identical scales are used to measure performance. Thus, the strength of the correlation between self-assessments of knowledge and cognitive learning should vary based on the degree of congruence in the self-assessment and cognitive learning measures.

Hypothesis 6: The relationship between self-assessments of knowledge and cognitive learning will be stronger when the constructs are assessed with similar rather than dissimilar measures.

Focus of Self-Assessment

Researchers differ in terms of whether they ask learners to self-assess their current knowledge levels or to assess the extent to which they have increased their knowledge (i.e., knowledge gain). For example, Carrell and Wilmington (1996) asked students in a communication course to rate their competence on six dimensions of interpersonal communication, an assessment of knowledge level. In contrast, Le Rouzie, Ouchi, and Zhou (1999) asked employees taking organizational training courses to rate the extent to which they acquired information that was new to them during training, an assessment of knowledge gain. Thus, focus of self-assessment indicates whether learners were asked to rate their knowledge level in the domain or how much knowledge they gained.

Asking learners to make absolute assessments of their knowledge levels is conceptually distinct from asking learners to rate their knowledge gain. An absolute assessment requires a judgment against an external standard, whereas asking learners about changes in their levels of knowledge requires self-referential judgments. It is possible for learners to be knowledgeable about a domain at the beginning of a course, learn little

from the course, and know the same amount when the course ends. These learners would have high levels of absolute knowledge and low levels of knowledge gain. Absolute self-assessments of knowledge are better matched with cognitive learning and should have stronger correlations with cognitive learning than self-assessments of knowledge gain.

Hypothesis 7: The relationship between self-assessments of knowledge and cognitive learning will be stronger when the self-assessment asks learners to report their current knowledge level than when the assessment asks learners to report their knowledge gain.

Self-Assessments in Evaluation Research and Practice

Researchers differ in their interpretations of self-assessments of knowledge. Some treat them as a facet of reactions (Kirkpatrick, 1998; Marsh, 1983), while others categorize them as an indicator of learners' knowledge levels (I-TECH, 2006). A cursory review of the literature does not yield a simple conclusion as to whether the interpretation of self-assessment has changed over time, but there are some obvious differences across research disciplines. A review of the communication literature, for example, revealed that self-assessments of knowledge are often used as an indicator of cognitive learning (e.g., Frymier, 1994; Hess & Smythe, 2001; Rodriguez, Plax, & Kearney, 1996). How self-assessments are used and interpreted in the business literature, however, is not entirely clear. Thus, we conducted a review of the self-assessment of knowledge literature to examine how self-assessment data are interpreted. Our review addresses three research questions: (1) How are self-assessments of knowledge used and interpreted? (2) Does the interpretation of self-assessments of knowledge differ across research disciplines? and (3) Has the interpretation of self-assessments changed over time?

METHOD

Sample of Studies

Computer-based literature searches of PsycInfo, ERIC, ProQuest, and Digital Dissertations were used to locate relevant studies. To be included in the initial review, studies had to contain terms relevant to self-assessments of knowledge and training or education. We used various forms of the

following keywords in our literature searches: (*training or education*) and (*perceive or self-report or self-assessment or self-evaluation*) and (*learn or performance or competence or knowledge*). Initial searches resulted in 12,718 possible studies. A review of abstracts limited the list to 573 potentially relevant studies, of which 77 contained one or more relevant effect sizes. In addition, we manually searched reference lists from a variety of published reports focusing on self-assessments of knowledge and all the papers included in the meta-analysis to identify relevant studies. These searches identified an additional 67 studies.

An extensive search for unpublished studies was also conducted. First, we manually searched the Academy of Management and Society for Industrial and Organizational Psychology conference programs from 1996 to 2007. Second, authors of studies already included in the meta-analysis, as well as other practitioners and researchers with expertise in training and education, were asked to provide leads on unpublished work. In all, we contacted 268 individuals. These efforts identified 22 additional studies for a total of 166 studies, yielding 222 independent samples.

Studies were included in the meta-analysis if (1) participants were nondisabled adults ages 18 or older, (2) the course facilitated potentially job-relevant knowledge or skills (i.e., not coping with physical or mental health challenges), and (3) relevant correlations were reported or could be calculated given the reported data. The first two criteria support generalization to adult education and training programs.

Coding and Interrater Agreement

Six potential moderators were coded for each study: external feedback on performance, delivery media, nature of the course content, practice self-assessing and feedback on the accuracy of self-assessments, similarity of measures, and focus of the self-assessment. External feedback on performance indicates whether learners received information on their performance during the course and was coded with two categories: no feedback or feedback was provided. Delivery media refers to whether the course was delivered with face-to-face classroom instruction, Web-based instruction, or blended learning (i.e., a combination of classroom and Web-based instruction). Nature of the course content was categorized as cognitive, interpersonal, or psychomotor. Cognitive content involves generating ideas, problem solving, and learning factual information (e.g., learning about the Central Limit Theorem or business principles),

whereas interpersonal content involves communication skills (e.g., giving a speech or articulating performance goals to one's team). Psychomotor content involves learning how to perform a skill (e.g., flying a plane or sketching designs for marketing material). Practice self-assessing and feedback on self-assessments was coded with three categories: learners did not practice self-assessing their knowledge, learners practiced self-assessing but did not receive feedback on the accuracy of their self-assessments, and learners practiced self-assessing and received feedback on the accuracy of their self-assessments. Similarity of measures indicates the correspondence between the level of specificity of the constructs measured in the self-assessment and cognitive learning domains. Studies were coded as using similar measures if the self-report and cognitive learning indicators were referring to the same criterion (e.g., how well will you perform on the first management exam with scores on that exam). Studies were coded as using dissimilar measures when the self-assessment and cognitive learning indicators were not matched in terms of specificity or the two assessments referred to different criteria (e.g., rate your ability to calculate an ANOVA with scores on a statistics test measuring a broad range of statistics knowledge). Finally, focus of self-assessment refers to whether the self-assessment of knowledge asked learners to report their knowledge gain (e.g., how much did you learn in this course?) or knowledge level (e.g., how knowledgeable are you about the material covered in this course?).

Four characteristics of the research reports were coded to address how self-assessments of knowledge are used in research: date, research discipline, use and interpretation of self-assessments of knowledge, and conclusion reached regarding the accuracy of self-assessments. Date was coded as the year of the publication, dissertation, or conference presentation. Research discipline was coded based on the journal in which the article was published, the department approving the dissertation, or the discipline affiliated with the conference presentation. Five categories were used to classify the use and interpretation of self-assessments of knowledge: (1) Studies were categorized as treating self-assessments as an indicator of learning if they referred to the self-assessment measure as learning or used the results to draw conclusions about learners' knowledge levels; (2) Self-assessments were categorized as reactions when the study referred to self-assessments as a facet of reactions, a measure of course satisfaction, or a level one evaluation criterion (Kirkpatrick, 1998); (3) Studies that used self-assessments as an evalua-

tion criterion but did not specifically state that the measure was tapping learning or reactions were categorized as interpreting self-assessments as a general evaluation criterion; (4) Studies that used self-assessments as an antecedent of evaluation outcomes were categorized as using self-assessments as predictors; (5) Studies that were conducted with the specific intent of determining the relationship between self-assessments and cognitive learning were categorized as examining the validity or accuracy of self-assessments. Conclusions reached regarding the accuracy of self-assessments were coded into three categories: inaccurate, mixed (i.e., there was variability across learners in the accuracy of their self-assessments), or accurate.

Three raters participated in the coding process. Each rater was given coding rules that specified the variables to code for the meta-analysis as well as specific definitions and examples for each coding category. They then attended a series of meetings where each rater independently coded an article, compared their coding, discussed coding discrepancies, and clarified the coding rules. After the raters developed a shared understanding of the coding rules, two raters independently coded each article in the meta-analysis, discussed discrepancies, and reached a consensus.

Interrater agreement (Cohen's kappa) was excellent according to Fleiss (1981) for each of the coded variables with a coefficient of .99 for date, .98 for both research discipline and delivery media, .92 for the use and interpretation of self-assessments of knowledge, .91 for both conclusion reached regarding the accuracy of self-assessments and nature of the course content, .90 for focus of the self-assessment, .88 for similarity of measures, .82 for external feedback on performance, and .79 for practice and feedback on self-assessments.

Independence Assumption

Multiple dimensions of reactions (e.g., affective and utility reactions) and posttraining motivation (e.g., motivation to transfer) were originally coded. However, single studies contributing multiple correlations can result in biased sampling error estimates (Hunter & Schmidt, 2004). Thus, when multiple measures of the same construct were present in a sample, the Hunter and Schmidt formula was used to calculate a single estimate that took into account the correlation among the measures. Studies that included multiple independent samples were coded separately and treated as independent.

Meta-Analytic Methods

The corrected mean and variance in validity coefficients across studies were calculated using formulas from Hunter and Schmidt (2004). The mean and variance of the correlations across studies were corrected for sampling error and unreliability in the predictor and criterion. Artifact distributions were created for each construct based on formulas from Hunter and Schmidt and are available upon request from the first author. Reliabilities for self-assessment, cognitive learning, and affective outcomes from all coded studies were included in the distributions. The average reliability was .86 for self-assessment, .71 for cognitive learning, .87 for reactions, .84 for motivation, and .85 for self-efficacy. When relevant, measures were corrected for dichotomization (e.g., when analyses reported were based on a median split of a continuous variable) using the formula provided by Hunter and Schmidt.

Prior to finalizing the analyses, a search for outliers was conducted using a modified Hufcutt and Arthur (1995) sample-adjusted meta-analytic deviancy (SAMD) statistic with the variance of the mean correlation calculated according to the formula specified by Beal, Corey, and Dunlap (2002). Based on the results of these analyses and inspection of the studies, no studies warranted exclusion.

Two indices were used to assess whether moderators may be operating. First, we used the 75% rule and concluded moderators may be operating when less than 75% of the variance was attributable to statistical artifacts. Second, credibility intervals were calculated using the corrected standard deviation around the mean corrected correlation. If the credibility interval was wide, we inferred that the mean corrected effect size was

probably the mean of several subpopulations and moderators may be operating (Whitener, 1990). Hunter and Schmidt's (2004) subgroup procedure was used to test for moderators when these criteria indicated moderators were possible. In addition, confidence intervals were calculated around the uncorrected correlations to determine whether meta-analytic correlations were significantly different from zero.

RESULTS

One-hundred sixty-six studies contributed data to the meta-analysis, including 112 published studies, 40 dissertations, and 14 unpublished studies. These studies reported 222 independent samples of data gathered from 41,237 learners. Learners were university students in 75% of studies, employees in 21% of studies, and military personnel in 4% of studies. Across all studies providing demographic data, the average age of learners was 31 years and 43% of participants were male.

Main Effect Results

Meta-analytic correlation results are presented in Table 1. Cohen's (1988) definition of effect sizes (small effect sizes are correlations of .10, moderate are .30, and large are .50) guided interpretation of the results. Hypothesis 1 predicts self-assessments of knowledge are more strongly related to affective than cognitive training outcomes. In support of Hypothesis 1, self-assessments of knowledge had a moderate mean corrected correlation with cognitive learning ($\rho = .34$) but large mean corrected correlations with reactions ($\rho = .51$) and motivation ($\rho = .59$). In both cases, the upper-bound of the 95%

TABLE 1
Meta-Analytic Correlations With Self-Assessments of Knowledge

	<i>k</i>	Total <i>N</i>	<i>N</i> Weighted Mean <i>r</i>	ρ	Sample Var (e) + Artifact Var (α)	Pop. Var	% Var due to Artifacts	95% Confidence Interval		80% Credibility Interval	
								Lower	Upper	Lower	Upper
Cognitive learning	137	16,951	.27	.34	.01	.07	15.30	.20	.33	-.01	.69
Reactions	105	28,612	.44	.51	.00	.07	5.71	.38	.49	.16	.85
Motivation	47	9,534	.50	.59	.00	.05	10.11	.41	.58	.30	.88
Self-efficacy	32	3,720	.37	.43	.01	.07	12.53	.24	.49	.09	.77

Note. *k* = the number of effect sizes included in the analysis; Total *N* = sum of the sample sizes of studies included in the analysis; *N* Weighted Mean *r* = sample size weighted mean *r*; ρ = mean correlation corrected for measurement error based on predictor and criterion reliabilities; Sample Var (e) + Artifact Var (α) = sampling error variance + variance due to differences in reliability in the predictor and criterion; Pop. Var = variance of the corrected correlations; % Var due to Artifacts = proportion of variance in the observed correlation due to statistical artifacts.

confidence interval for cognitive learning ($r = .33$) was lower than the lower-bound of the 95% confidence interval for reactions ($r = .38$) and motivation ($r = .41$). However, the mean corrected correlation between self-assessments of knowledge and self-efficacy was .43, and the 95% confidence interval for this relationship overlaps with the confidence interval for the relationship between self-assessments of knowledge and cognitive learning. Thus, the trend supports Hypothesis 1, but the findings do not hold for all three affective outcome measures.

Moderator Analysis Results

The main effect analysis for self-assessments of knowledge and cognitive learning had a wide 80%

credibility interval (width was .70), and the percent of variance attributed to statistical artifacts (i.e., sampling error and between study differences in reliability) was less than 75%. Together the credibility interval and 75% rule suggest a high probability of multiple population values (Hunter & Schmidt, 2004). Thus, it is appropriate to test for moderators.

Hunter and Schmidt's (2004) subgroup analysis was used to test for moderators. Meta-analytic estimates for each subgroup are presented in Table 2. We concluded that the moderator had a meaningful effect when the average corrected correlation varied between subgroups by at least .10 and at least one of the estimated population variances was less than the estimated population variance for the entire sample of studies. This is consistent

TABLE 2
Moderator Analyses of the Relationship Between Self-Assessments of Knowledge and Cognitive Learning

	<i>k</i>	Total <i>N</i>	<i>N</i> Weighted Mean <i>r</i>	ρ	Sample Var (e) + Artifact Var (a)	Pop. Var	% Var due to Artifacts	95% Confidence Interval		80% Credibility Interval	
								Lower	Upper	Lower	Upper
Externally Generated Feedback											
Feedback – no	22	2,609	.11	.14	.01	.12	10.48	–.06	.28	–.31	.59
Feedback – yes	49	4,703	.21	.28	.01	.06	23.10	.13	.30	–.03	.58
Delivery Medium											
Classroom instruction	86	9,048	.25	.33	.01	.07	18.15	.17	.34	–.02	.34
Blended learning	7	425	.27	.34	.01	.09	22.28	–.02	.56	–.04	.67
Web-based instruction	12	1,878	.15	.19	.01	.01	46.47	–.03	.33	.05	.73
Nature of Course Content											
Cognitive	64	7,467	.25	.33	.01	.08	15.05	.17	.34	–.04	.69
Interpersonal	25	3,223	.41	.52	.01	.08	13.86	.26	.55	.17	.88
Psychomotor	15	528	.12	.15	.03	.03	62.88	–.13	.37	–.07	.38
Number of Self-Assessments During Training and Feedback on Accuracy											
One self-assessment	87	11,482	.23	.29	.01	.07	14.72	.14	.31	–.06	.64
Two or more self-assessments and no feedback on accuracy	34	2,968	.23	.30	.01	.03	39.53	.14	.33	.09	.52
Two or more self-assessments and feedback provided on accuracy	10	1,168	.40	.51	.01	.04	24.78	.21	.59	.25	.77
Similarity Between Self-Assessment and Cognitive Learning Measures											
Similarity – low	79	10,261	.19	.24	.01	.08	14.20	.12	.25	–.12	.60
Similarity – high	73	7,633	.36	.47	.01	.04	30.00	.27	.45	.23	.71
Focus of Self-Assessment Measure											
Knowledge gain	25	3,235	.00	.00	.01	.04	24.17	–.12	.12	–.26	.27
Knowledge level	108	12,296	.34	.44	.01	.06	22.40	.27	.41	.16	.72

Note. *k* = the number of effect sizes included in the analysis; Total *N* = sum of the sample sizes of studies included in the analysis; *N* Weighted Mean *r* = sample size weighted mean *r*; ρ = mean correlation corrected for measurement error based on predictor and criterion reliabilities; Sample Var (e) + Artifact Var (a) = sampling error variance + variance due to differences in reliability in the predictor and criterion; Pop. Var = variance of the corrected correlations; % Var due to Artifacts = proportion of variance in the observed correlation due to statistical artifacts.

with the criteria for moderation used by Sitzmann et al. (2008).

Hypothesis 2 predicts the relationship between self-assessments and cognitive learning is stronger in courses that provide external feedback on learners' performance than in courses that do not provide feedback. Supporting Hypothesis 2, the correlation was stronger for courses that included feedback ($\rho = .28$) than for courses that did not include feedback ($\rho = .14$). For this moderator result and all subsequent moderator results, at least one of the subgroup variances was lower than the overall population variance.

Hypothesis 3 predicts the relationship between self-assessments and cognitive learning is stronger in classroom instruction and blended learning than in Web-based instruction. The correlation was stronger for classroom instruction ($\rho = .33$) and blended learning ($\rho = .34$) than Web-based instruction ($\rho = .19$), supporting Hypothesis 3.

Hypothesis 4 predicts the relationship between self-assessments and cognitive learning is stronger in courses targeting interpersonal and psychomotor skills than in courses targeting cognitive skills. The correlation was stronger for interpersonal ($\rho = .52$) than cognitive ($\rho = .33$) courses. However, in psychomotor courses, self-assessments and cognitive learning were weakly related ($\rho = .15$). Thus, the results partially support Hypothesis 4.

Hypothesis 5 predicts the relationship between self-assessments of knowledge and cognitive learning is stronger in courses where learners practice making self-assessments and receive feedback on their accuracy. In courses where learners self-assessed their knowledge once, the corrected correlation between self-assessments and learning was .29. A similar relationship was observed when learners rated their knowledge multiple times, but did not receive feedback on the accuracy of their self-assessments ($\rho = .30$). However, the self-assessment-cognitive learning relationship was much stronger in courses where learners practiced self-assessing their knowledge and received feedback on their accuracy ($\rho = .51$), supporting Hypothesis 5.

Hypothesis 6 predicts the relationship between self-assessments and cognitive learning is stronger when similar measures are used to assess the constructs. In support of Hypothesis 6, when the measures were similar, self-assessments exhibited a stronger correlation with cognitive learning ($\rho = .47$) than when the measures were less similar ($\rho = .24$).

Hypothesis 7 predicts the relationship between self-assessments and cognitive learning is stronger when the focus of the self-assessment is on

knowledge levels rather than knowledge gains. When the focus of the self-assessment was learners' knowledge level, self-assessments of knowledge had a stronger correlation with cognitive learning ($\rho = .44$) than when the focus of the self-assessment was knowledge gain ($\rho = .00$).¹ Thus, Hypothesis 7 is supported.

Overall, the moderator results indicate the relationship between self-assessments and cognitive learning is strongest when (1) learners receive external feedback, (2) the delivery medium is classroom or blended instruction, (3) the course teaches interpersonal skills, (4) learners practice and receive feedback on the accuracy of their assessments, (5) similar measures are used to assess the two constructs, and (6) the focus of the self-assessment is learners' knowledge level.

A limitation of the subgroup approach for examining moderators is that it is restricted to testing individual hypotheses and does not account for possible confounds between correlated moderators (Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Miller, Glick, Wang, & Huber, 1991). To address this concern, we calculated the meta-analytic corrected correlation for studies that met at least five out of six moderator conditions associated with a stronger relationship between self-assessments and cognitive learning. In this analysis, self-assessments were strongly related to cognitive learning ($\rho = .55$, $k = 9$, $N = 937$), and the population variance was much smaller than the variance in the main effect ($\sigma^2 = .02$ vs. $.07$, respectively). This suggests that the effect of the moderators is additive, and variability in the self-assessment-cognitive learning relationship is greatly reduced by accounting for course design and methodological factors.

Interpretation of Self-Assessments of Knowledge

We reviewed the literature to examine how self-assessments of knowledge are interpreted. The results suggest 32% of studies interpreted self-assessments of knowledge as an indicator of learning (see Table 3). For example, Alavi (1994) used a self-assessment instrument to measure MBA students' learning in management information systems. An additional 7% of research reports interpreted self-assessments as an indicator of reactions (e.g., Dixon, 1990, included self-assessment

¹ Controlling for pretraining knowledge with meta-analytic regression analysis and correlating self-assessment gains with cognitive learning residuals (to represent a commensurate measure of gain) did not strengthen this relationship. These analyses are available upon request from the first author.

TABLE 3
Frequency Analysis of Research Discipline by the Use and Interpretation
of Self-Assessments of Knowledge

	Learning	Reactions	Criterion General	Predictor of Learning Outcomes	Examine the Validity or Accuracy of Self- Assessments
Psychology	22% (11)	6% (3)	33% (17)	14% (7)	25% (13)
Education	27% (10)	11% (4)	30% (11)	8% (3)	24% (9)
Communication	79% (19)	0% (0)	4% (1)	0% (0)	17% (4)
Business	30.5% (7)	22% (5)	30.5% (7)	4% (1)	13% (3)
Medical education	17% (2)	0% (0)	8% (1)	25% (3)	50% (6)
Foreign language	18% (2)	0% (0)	36% (4)	0 (0)	46% (5)
Other disciplines	25% (2)	0% (0)	25% (2)	0% (0)	50% (4)
Total	32% (53)	7% (12)	26% (43)	8% (14)	27% (44)

Note. The first number is the percent of studies within the discipline. The number in parentheses is the number of studies. Total number of studies is 166.

of knowledge as one facet of a reactions survey), whereas 26% interpreted self-assessments as a general evaluation criterion (e.g., Wise, Chang, Duffy, & Del Valle, 2004, included self-assessment as one component of a broad course evaluation). Finally, 8% of studies used self-assessment data as a predictor of learning outcomes, whereas 27% were conducted to determine the validity or accuracy of self-assessments.

There was a significant difference in the tendency to interpret self-assessments of knowledge as an indicator of learning across research disciplines ($\chi^2[6, N = 166] = 21.90, p < .05$). In communication, self-assessments were interpreted as an indicator of learning in 79% of reports. Business was the second most likely discipline to interpret self-assessments as an indicator of learning (30.5%), followed by education (27%), psychology (22%), foreign language (18%), and medical education (17%).²

Next, we focused on the studies that drew conclusions regarding the accuracy of self-assessments to determine what conclusions were reached (Table 4). Overall, 56% of the 55 relevant studies concluded that self-assessments of knowledge were inaccurate, whereas 26% concluded learners were accurate, and 18% reported mixed results. There was not a significant difference across research disciplines in the conclusions reached regarding the accuracy of self-assessments of knowledge ($\chi^2[6, N = 55] = 7.37, p > .05$), but this may have been due to the small sample size and associated low statistical power. In business studies,

researchers never concluded that learners accurately self-assessed their knowledge, whereas 80% concluded they were inaccurate and 20% found mixed results. On the other extreme, 60% of foreign language studies and 55% of education studies concluded learners were accurate.

We also examined whether the tendency to interpret self-assessments as an indicator of learning has changed over time (Table 5). The results suggest there was not a significant difference in the percent of studies that interpreted self-assessments as a learning indicator in articles published from 1954 to 1989, the 1990s, or the 2000s ($\chi^2[2, N = 108] = 3.19, p > .05$). However, fewer

TABLE 4
Frequency Analysis of the Conclusions Reached
Regarding the Accuracy of Self-Assessments of
Knowledge by Discipline

	Conclusion Regarding the Accuracy of Self-Assessments of Knowledge		
	Inaccurate	Mixed Results	Accurate
Psychology	40% (6)	47% (7)	13% (2)
Education	45% (5)	0% (0)	55% (6)
Communication	86% (6)	0% (0)	14% (1)
Business	80% (4)	20% (1)	0% (0)
Medical education	62.5% (5)	25% (2)	12.5% (1)
Foreign language	40% (2)	0% (0)	60% (3)
Other disciplines	75% (3)	0% (0)	25% (1)
Total	56% (31)	18% (10)	26% (14)

Note. The first number is the percent of studies within the discipline. The number in parentheses is the number of studies. All studies that reached a conclusion regarding the accuracy of self-assessments were included in the analysis, regardless of whether the purpose of the study was to examine the accuracy of self-assessments. Total number of studies is 55.

² In follow-up analyses, we attempted to divide business and the other disciplines into more precise subdisciplines. However, the small sample sizes made these analyses difficult to interpret.

TABLE 5
Frequency Analysis of the Interpretation of Self-
Assessments of Knowledge as an Indicator of
Learning by Decade

	1980s and Earlier	1990s	2000s
Not interpreted as an indicator of learning	67% (4)	40% (17)	57% (34)
Interpreted as an indicator of learning	33% (2)	60% (25)	43% (26)

Note. The first number is the percent of studies within the decade. The number in parentheses is the number of studies. We focused exclusively on studies that were using self-assessments as an evaluation criterion. Total number of studies is 108.

studies interpreted self-assessments as an indicator of learning in the 2000s (43% of 60 studies) than in the 1990s (60% of 42 studies). In articles published before 1990, 33% (of 6 studies) interpreted the data as an indicator of learning.

DISCUSSION

A meta-analysis was conducted to examine the construct validity of self-assessments of knowledge. Specifically, we analyzed the relationships between self-assessments of knowledge and cognitive learning, reactions, motivation, and self-efficacy. We also tested course design and measurement characteristics as moderators of the validity of self-assessments of knowledge. Finally, we examined how self-assessments are used and interpreted in course evaluations. In the following sections, we discuss both theoretical and practical implications of our results, study limitations, and directions for future research.

Theoretical Implications

The meta-analytic results revealed self-assessments are best categorized as an affective evaluation outcome. Self-assessments of knowledge were strongly related to reactions and motivation and moderately related to self-efficacy. The corrected mean correlations between these constructs and self-assessed knowledge were also greater in magnitude than the self-assessment-cognitive learning relationship. Even in evaluation contexts designed to optimize the self-assessment-learning relationship (e.g., when similar measures were used and feedback was provided), self-assessments had as strong of a relationship with motivation as with cognitive learning. These results suggest that self-assessed knowledge is generally more useful as

an indicator of how learners feel about a course than as an indicator of how much they learned from it.

As noted earlier, these results are consistent with concerns advanced in prior research regarding self-assessment. Dunning et al. (2004) suggested that accurately self-assessing one's performance is intrinsically difficult. Learners must have a clear understanding of the performance domain in question. They also must be willing to consider all aspects of their knowledge levels (not just the favorable) and to overcome the egocentrism that results in people assuming they are above average in most aspects of their lives. This is difficult for incompetent learners, given that they lack the insight required to both self-assess their knowledge and understand the material. These results are also consistent with trends moving away from self-assessments as an indicator of learning. Most notably, AACSB International accreditation standards now suggest their member institutions should rely on tests or rated performance rather than self-assessments as assurance of learning.

The vast majority of research included in the meta-analysis used self-assessments as part of summative course evaluations. *Summative evaluations* refer to end-of-course evaluations used to determine its effectiveness or efficiency. Summative evaluations are often contrasted with *formative evaluations*, which refer to evaluations that occur during a course and are aimed at improving the effectiveness or efficiency of the course as it is being developed or delivered. The dominance of data from summative evaluation efforts is consistent with the current state of evaluation research in psychology, management, and related fields, which overwhelmingly focus on summative evaluation (Brown & Gerhardt, 2002). However, some researchers have suggested that self-assessments are more useful—or are only useful—when measured as part of a formative evaluation (e.g., Arthur, 1999; Fuhrmann & Weissburg, 1978). Including self-assessments as part of a formative evaluation encourages learners to identify their strengths and weaknesses (Stuart, Goldstein, & Snope, 1980) and may stimulate interest in learning and professional development (Arnold, Wiloughby, & Calkins, 1985).

We believe self-assessments are an important component of the learning process and should be understood in terms of how they influence where learners direct their study time and energy. Drawing from control theory (Carver & Scheier, 1990), learners entering a course engage in a variety of self-regulatory activities that include developing cognitive strategies and applying effort in an at-

tempt to learn the material. Periodically they self-assess their learning progress, which leads to deciding whether there is a goal-performance discrepancy—a discernable difference between desired and actual knowledge (Carver & Scheier, 2000). When learners detect a discrepancy, it influences their learning behaviors (Campbell & Lee, 1988; Carver & Scheier, 1990; Winne, 1995; Zimmerman, 1990). If learners believe they have not met their goals, they will continue to develop strategies and apply effort (Carver & Scheier, 1990, 2000). A variety of other theories also suggest that learners use their self-assessments to modify their study strategies, regulate effort levels, and focus their cognitive resources on unfamiliar material, which should facilitate learning (e.g., Kanfer & Kanfer, 1991; Pintrich, 1990; Winne, 1995). Our research suggests that self-assessments are strongly related to actual knowledge levels when learners are given an opportunity to self-assess and receive feedback on their self-assessments. This is consistent with prior research that emphasizes self-assessment as a valuable skill. Thus, we encourage further research on the role of self-assessments in the learning process and suggest that self-assessments should be administered throughout the course with the objective of understanding how quickly they improve in accuracy and how they influence learning processes.

Self-Assessments in Evaluation Research

A significant number of research studies interpret self-assessments of knowledge as indicators of learning. Even in more recent research (2000 to present), 43% of studies interpreted self-assessment data as evidence of learning. Because self-assessments do not always correlate with learning, this suggests that there may be a disconnect between validity evidence and the interpretation of self-assessments. This potential disconnect was most prevalent in the communication literature. For example, Chesebro and McCroskey (2000) reported a correlation of .50 between self-assessments and knowledge among 192 learners. Their report concluded with the claim that “the results of this study are indeed useful, because they support the notion that students can report accurately on their own learning” (301). Despite more recent evidence in the communication domain indicating that self-assessments are sometimes weak and imperfect indicators of learning (Hess & Smythe, 2001; Witt & Wheelless, 2001), researchers have continued citing Chesebro and McCroskey to justify the use of self-assessments in lieu of more objective knowledge measures (e.g., Benbunan-Fich & Hiltz, 2003; Teven, 2005).

Communication is not the only discipline that interprets self-assessments as indicators of cognitive learning. In the business domain, 30.5% of reports used self-assessed knowledge as an indicator of learning. This occurred despite the fact that in this literature 80% of studies that evaluated the accuracy of self-assessments concluded that learners’ self-assessments were inaccurate. This finding is consistent with research indicating there is extensive evidence of a science–practice gap in human resource management (e.g., Rynes, Colbert, & Brown, 2002). As such, scientific findings regarding the construct validity of self-assessments may not have influenced the interpretation of self-assessment data. The causes for this gap are numerous, but include such factors as failure to read research findings (Rynes et al., 2002) and failure to implement such findings because of situational forces (Pfeffer & Sutton, 1999).

Practical Implications

Our results have clear implications for the design and evaluation of courses as well as for needs assessment practices. In terms of instructional design, research suggests that self-assessments of knowledge have a critical role in the learning process and that learners benefit from having an accurate understanding of their knowledge levels (Dunning et al., 2004). Specifically, in order for learners to build lifelong learning habits, they must be able to critically evaluate their own knowledge and skill levels (Sullivan, Hitchcock, & Dunnington, 1999). Thus, management courses should be designed to develop learners’ self-assessment skills and promote a strong correspondence between learners’ self-assessments and their knowledge levels.

We provide two recommendations to assist instructional designers in strengthening the relationship between learners’ self-assessments and their knowledge levels. First, learners should be provided with periodic feedback on their performance in the course. Feedback provides learners with descriptive information about their performance (Bandura, 1977; Bell & Kozlowski, 2002) and can assist them in calibrating their self-assessments of knowledge. Second, learners should have the opportunity to practice self-assessing their knowledge and to receive feedback on the accuracy of their self-assessments. Self-assessment may be a skill that learners can acquire by way of practice and feedback on their ratings (Levine et al., 1977). As such, we recommend that management courses and curricula be developed with the goal of fostering self-assessment skills.

In terms of evaluating management courses, it is our hope that this meta-analysis will lead researchers and practitioners to be more prudent in their use of self-assessments of knowledge as a cognitive learning measure. Self-assessments seem to be influenced by how learners felt about their course experience and by their motivation levels. Thus, their self-reports of knowledge may be systematically biased such that interesting and fun learning experiences also receive favorable self-assessment ratings.

Although self-assessments exhibited moderate to strong relationships with reactions, motivation, and self-efficacy, we would not recommend using self-assessed learning as an indicator of these constructs. If these constructs are of theoretical interest, then they should be measured directly. For example, if the practical question being considered is whether learners will take follow-up courses, then affective outcomes, such as reactions and motivation, are sensible. However, if the question is whether learners gain more knowledge using a particular teaching approach, then knowledge tests should be incorporated in courses. This particular approach to evaluation is consistent with the recommendations of Kraiger (2002) and Brown (2006), who note that the purpose of the evaluation should determine which measures are included in the evaluation effort.

We acknowledge that using tests rather than short, self-report measures can increase the cost and time required for evaluations. Administrators, professors, managers, and even professional evaluators seeking to reduce the time spent developing and administering evaluation instruments often prefer a short, easy-to-administer measure. Self-assessments are also appealing because they can be used across courses, circumventing the need to develop commensurate tests (e.g., Arbaugh & Durray, 2002). However, in these situations, we recommend an approach similar to Arbaugh (2005)—using knowledge measures, such as course grades, that are adjusted to control for instructor-level effects. Alternatively, performance artifacts, such as written work, can be rated using a common rubric.

As a final implication, we believe these findings support prior calls to include direct indicators of knowledge in organizations' needs assessment efforts (Tracey, Arroll, Barham, & Richmond, 1997). If self-assessments are only moderately related to actual knowledge levels, then basing learning needs and subsequent course-related decisions on these ratings may lead organizations to overlook areas of knowledge deficiency. Although inconvenient, short knowledge tests made available on-

line might be a useful way to determine the topics on which employees need training.

Study Limitations and Directions for Future Research

There are several limitations to our study. First, our meta-analysis focused on studies that reported correlations between self-assessments of knowledge and other course evaluation criteria (i.e., cognitive learning, reactions, motivation, and self-efficacy). Thus, we do not fully capture how self-assessments are used in research, and we probably understate the degree to which self-assessments are used as an indicator of learning. It is possible that researchers using self-assessments as their only course evaluation criterion may be even stronger advocates of the value of these measures than the studies included in our review.

Second, only 23 studies (14%) included in the meta-analysis fell within in the business discipline. This suggests that additional research on the validity of self-assessments of knowledge may be warranted within the business domain. Moreover, business is not a homogenous discipline, and it is possible that the self-assessment–cognitive learning relationship may differ across business subdisciplines. The limited number of studies that have examined the validity of self-assessments of knowledge within business made it challenging to draw conclusions about subdisciplines, but this is an important avenue for future research.

Third, none of the moderator variables independently accounted for all of the variability in the self-assessment–learning relationship, and the results of the compound moderator analysis suggest that there are moderators of this relationship that were not identified here. Thus, additional research is needed to investigate learner and course design characteristics that moderate the self-assessment–learning relationship. For example, Kruger and Dunning (1999) found learners' level of expertise in the domain influenced their self-assessments. We were unable to code for this as a between-study moderator because few studies provided sufficient detail to indicate learners' levels of prior knowledge.

Another moderator that we were unable to code for is whether the self-assessment was made for developmental or administrative purposes. Recent research in the performance appraisal domain has shown that rater characteristics and the context of the rating influence individual's self-assessments of job performance. For example, Patiar and Mia (2008) found that women tended to rate themselves lower than their male counterparts, and research suggests that self-ratings made for developmental

purposes are more accurate than ratings made for administrative purposes (Zimmerman, Mount, & Goff, 2008). In terms of course evaluations, self-assessments may be used to examine whether the instructor should spend additional time covering a topic in a course or as a component of student grades. It is possible that self-assessments used for developmental purposes are more accurate than those used for grading students. Moreover, research should examine the effect of learner characteristics on the relationship between self-assessments and cognitive learning.

Fourth, most studies measured self-assessment and evaluation outcomes at a single point in time and provided only limited descriptions of the course. This prevented us from examining more specific information about courses that might influence whether the validity of self-assessments changes as learners progress through the course. The influence of additional course design characteristics on the validity of self-assessments and on changes in validity over time would be useful avenues for future research. For example, evaluation research emphasizes the importance of providing feedback to improve learning (e.g., Sitzmann, Kraiger, Stewart, & Wisher, 2006), and the current results suggest feedback promotes self-assessments that are moderately to strongly related to cognitive learning. Thus, it is possible that feedback influences learning by way of promoting accurate self-assessments that enable learners to focus their cognitive resources and apply effort toward learning the material they have not mastered. Future research should model changes in learning as learners progress through a course and manipulate which modules provide feedback. If the correspondence between self-assessments and cognitive learning covaries positively with whether feedback was provided, it suggests feedback plays a vital role in the self-assessment–learning relationship.

Fifth, results from the current meta-analysis suggest self-assessments and learning are only moderately correlated. However, the current study does not provide a definitive reason as to why the relationship is not stronger. Research from the performance appraisal literature suggests there are several reasons why subjective assessments and actual performance may not be highly correlated (Murphy, 2008). One possibility is that individuals may not be skilled at making ratings (Murphy). Kruger and Dunning (1999) found that poor performers lack the metacognitive skills necessary to self-assess their knowledge. Another reason is that individuals may allow other factors, such as individual differences, to influence their ratings (Murphy). For example, trainees with a high per-

formance-prove goal orientation may let their desire to do better than other trainees influence their ratings, causing them to inflate their self-assessments of knowledge. Future research is needed to parse these issues to better understand the relationship between self-assessments and learning.

Sixth, the vast majority of studies in this meta-analysis were conducted with learners from the United States. Research has demonstrated that people from Western or individualistic cultures exhibit a *leniency bias*, the tendency to rate oneself more favorably than others would (Farh & Werbel, 1986; Holzbach, 1978; Nilsen & Campbell, 1993). However, individuals from Eastern or collectivistic cultures have been shown to exhibit a *modesty bias*, where they underrate their performance (Farh, Dobbins, & Cheng, 1991; Yik, Bond, & Paulhus, 1998). Thus, the results of this meta-analysis may not generalize to collectivistic cultures.

Understanding the construct validity of self-assessments may also provide insight as to how learners allocate their time in learner-controlled study environments. Research suggests some learners who are provided with control over their learning experience focus on material they have already mastered and exit the course before mastering all of the course content (Bell & Kozlowski, 2002; Brown, 2001). Thus, research is needed that models changes in self-assessments of knowledge as learners progress through learner-controlled courses. Specifically, do self-assessments of knowledge predict time spent in the course? Are learners with inflated self-assessments likely to exit the course before mastering its content, or is time spent in the course better predicted by individual differences, such as conscientiousness?

Finally, research is needed to examine the effect of self-regulatory processes on self-assessments of knowledge. Are learners who develop metacognitive strategies for learning the material more likely to have accurate self-assessments of knowledge? Do learners experience negative affect when their self-assessments of knowledge indicate their progress toward their goals is slower than expected? Explicitly measuring metacognition, affect, and self-assessments of knowledge would allow us to test the assumptions of Carver and Scheier's (1990) control theory and aid our understanding of how self-assessments influence learning processes.

CONCLUSION

This research clarifies that self-assessments of knowledge are only moderately related to cognitive learning and are strongly related to affective

evaluation outcomes. Even in evaluation contexts that optimized the self-assessment-learning relationship (e.g., when learners practiced self-assessing and received feedback on their self-assessments), self-assessments had as strong of a relationship with motivation as with cognitive learning. Thus, we encourage researchers to be prudent in their use of self-assessed knowledge as a cognitive learning outcome, taking into account characteristics of the learning environment before making claims about the usefulness of the self-assessment measure. We also encourage future research on the self-assessment process, and more specifically, how educators and trainers can build accurate self-assessments that promote lifelong learning.

REFERENCES

- Note: References marked with an asterisk indicate studies included in the meta-analysis.
- AACSB International 2007. **AACSB Assurance of Learning Standards: An interpretation**. Available on-line at http://www.aacsb.edu/resource_centers/assessment.
- *Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D. J., & Romero, V. L. 2000. Evaluation of an interactive tutorial for teaching the Central Limit Theorem. *Teaching of Psychology*, 27: 289–291.
- *Abramowitz, S. K. 1999. A computer-enhanced course in descriptive statistics at the college level: An analysis of student achievement and attitudes. Unpublished doctoral dissertation, New York University.
- *Al-Ammar, S. 1994. The influence of individual and organizational characteristics on training motivation and effectiveness. Unpublished doctoral dissertation, University at Albany, NY.
- *Alavi, M. 1994. Computer-mediated collaborative learning: An empirical evaluation. *MIS Quarterly*, 18: 159–174.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. 1997. A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50: 341–358.
- *Almegta, N. R. 1996. *Relationship of self-efficacy, causal attribution, and emotions to female college students' academic self-evaluation*. Unpublished doctoral dissertation, The University of Arizona, Tucson.
- Ambady, N., & Rosenthal, R. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64: 431–441.
- *Arbaugh, J. B. 2005. How much does "subject matter" matter? A study of disciplinary effects in on-line MBA courses. *Academy of Management Learning and Education*, 4: 57–73.
- *Arbaugh, J. B., & Duray, R. 2002. Technological and structural characteristics, student learning and satisfaction with Web-based courses: An exploratory study of two on-line MBA programs. *Management Learning*, 33: 331–347.
- *Arnold, H. J. 1985. Task performance, perceived competence, and attributed causes of performance as determinants of intrinsic motivation. *Academy of Management Journal*, 28: 876–888.
- *Arnold, L., Willoughby, T. L., & Calkins, E. V. 1985. Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education*, 60: 21–28.
- Arreola, R. 2006. *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system* (3rd ed.). Bolton, MA: Anker Publishing Company.
- Arthur, H. 1999. Student self-evaluation: How useful? How valid? *International Journal of Nursing Studies*, 32: 271–276.
- Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. 2003. Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, 88: 234–245.
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. 1995. The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, 29: 432–442.
- *Baird, J. S. 1987. Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology*, 79: 90–91.
- *Baker, J. D. 2001. The effects of instructor immediacy and student cohesiveness on affective and cognitive learning in the online classroom. Unpublished doctoral dissertation, Regent University, Virginia Beach, VA.
- *Bakx, A. W. E. A., van der Sanden, J. M. M., Sijtsma, K., Croon, M. A., & Vermetten, Y. J. M. (n.d.). Acquiring social-communicative competence: a developmental model for domain-related learning. Unpublished manuscript.
- *Bakx, A. W. E. A., van der Sanden, J. M. M., Sijtsma, K., Croon, M. A., & Vermetten, Y. J. M. 2006. The role of students' personality characteristics, self-perceived competence and learning conceptions in the acquisition and development of social communicative competence: A longitudinal study. *Higher Education*, 51, 71–104.
- Bandura, A. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- *Barton-Arwood, S., Morrow, L., Lane, K., & Jolivette, K. 2005. Project IMPROVE: Improving teachers' ability to address students' social needs. *Educational and Treatment of Children*, 28, 430–443.
- Beal, D. J., Corey, D. M., & Dunlap, W. P. 2002. On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in meta-analysis of correlations. *Journal of Applied Psychology*, 87, 583–589.
- *Bell, B. S., & Ford, J. K. 2007. Reactions to skill assessment: The forgotten factor in explaining motivation to learn. *Human Resource Development Quarterly*, 18, 33–62.
- Bell, B. S., & Kozlowski, S. W. J. 2002. Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, 55, 267–307.
- *Bell, B. S., & Roberson, Q. M. 2006. [Diversity training research.] Unpublished raw data.
- *Benbunan-Fich, R., & Hiltz, S. R. 2003. Mediators of the effectiveness of online courses. *IEEE Transactions on Professional Communication*, 46, 298–312.
- *Bergee, M. J. 1997. Relationships among faculty, peer, and self-

- evaluations of applied performances. *Journal of Research in Music Education*, 45, 601–612.
- *Bergee, M. J., & Cecconi-Roberts, L. 2002. Effects of small-group peer interaction on self-evaluation of music performance. *Journal of Research in Music Education*, 50, 256–268.
- *Bishop, J. B. 1971. Another look at counselor, client, and supervisor ratings of counselor effectiveness. *Counselor Education and Supervision*, 10, 319–323.
- *Boud, D. J., & Tyree, A. L. 1980. Self and peer assessment in professional education: A preliminary study in law. *Journal of the Society of Public Teachers of Law*, 15, 65–74.
- Brown, K. G. 2001. Using computers to deliver training: Which employees learn and why? *Personnel Psychology*, 54, 271–296.
- Brown, K. G. 2005. An examination of the structure and nomological network of trainee reactions: A closer look at "smile sheets." *Journal of Applied Psychology*, 90, 991–1001.
- Brown, K. G. 2006. Training evaluation. In S. Rogelberg (Ed.), *The Encyclopedia of Industrial and Organizational Psychology*, 2: 820–823. Thousand Oaks, CA: Sage.
- Brown, K. G., & Gerhardt, M. W. 2002. Formative evaluation: An integrative practice model and case study. *Personnel Psychology*, 55: 951–983.
- *Burke, L. A. 1997. Improving positive transfer: A test of relapse prevention training on transfer outcomes. *Human Resource Development Quarterly*, 8: 115–128.
- *Butler, A., Phillman, K. B., & Smart, L. 2001. Active learning within a lecture: Assessing the impact of short, in-class writing exercises. *Teaching of Psychology*, 28: 257–259.
- Butler, D. L., & Winne, P. H. 1995. Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65: 245–281.
- Campbell, D. J., & Lee, C. 1988. Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review*, 13: 302–314.
- *Carpenter, V. L., Friar, S., & Lipe, M. 1993. Evidence on the performance of accounting students: Race, gender and expectations. *Issues in Accounting Education*, 8: 1–17.
- *Carrell, L. J., & Willmington, S. C. 1996. A comparison of self-report and performance data in assessing speaking and listening competence. *Communication Reports*, 9: 185–191.
- *Carswell, A. D. 2001. *Facilitating student learning in an asynchronous learning network*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Carver, C. S., & Scheier, M. F. 1990. Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97: 19–35.
- Carver, C. S., & Scheier, M. F. 2000. Scaling back goals and recalibration of the affect system are processes in normal adaptive self-regulation: Understanding "response shift" phenomena. *Social Science & Medicine*, 50: 1715–1722.
- Centra, J. A., & Gaubatz, N. B. 1998. Is there gender bias in student ratings of instruction? *Journal of Higher Education*, 70: 17–33.
- *Chemers, M. M., Hu, L., & Garcia, B. F. 2001. Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology*, 93: 55–64.
- *Chesebro, J. L., & McCroskey, J. C. 2000. The relationship between students' reports of learning and their actual recall of lecture material: A validity test. *Communication Education*, 49: 297–301.
- *Chesebro, J. L., & McCroskey, J. C. 2001. The relationship of teacher clarity and immediacy with student state receiver apprehension, affect, and cognitive learning. *Communication Education*, 50: 59–68.
- *Chiaburu, D. S., & Marinova, S. V. 2005. What predicts skill transfer? An exploratory study of goal orientation, training self-efficacy and organizational supports. *International Journal of Training and Development*, 9: 110–123.
- *Chiaburu, D. S., & Tekleab, A. G. 2005. Individual and contextual influences on multiple dimensions of training effectiveness. *Journal of European Industrial Training*, 29: 604–626.
- *Christophel, D. M. 1990. The relationship among teacher immediacy behaviors, student motivation and learning. *Communication Education*, 39: 323–340.
- *Chur-Hansen, A. 2000. Medical students' essay-writing skills: Criteria-based self- and tutor-evaluation and the role of language background. *Medical Education*, 34: 194–198.
- *Clayson, D. E. 2005. Performance overconfidence: Metacognitive effects or misplaced student expectations? *Journal of Marketing Education*, 27: 122–129.
- Cohen, P. A. 1981. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51: 281–309.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*: 2nd ed. Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., & Meehl, P. E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- *Cuevas, H. M. 2004. Transforming learning into a constructive cognitive and metacognitive activity: Use of a guided learner-generated instructional strategy within computer-based training. Unpublished doctoral dissertation, University of Central Florida, Orlando.
- *Curry, D. H. 1997. Factors affecting the perceived transfer of learning of child protection social workers. (Doctoral dissertation, Kent State University, 1997). *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 58: 735.
- d'Apollonia, S., & Abrami, P. C. 1997. Navigating student ratings of instruction. *American Psychologist*, 52: 1198–1208.
- *Dark, M. J. 2000. Exploring systematic relationships of evaluation data for a reduced model of evaluation. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.
- Darwin, C. 1871. *The descent of man*. London: John Murray.
- *Das, M., Mpofu, D., Dunn, E., & Lanphear, J. H. 1998. Self and tutor evaluation in problem-based learning tutorials: Is there a relationship? *Medical Education*, 32: 411–418.
- *Davis, W. D. 1998. A longitudinal field investigation of antecedents and consequences of self-efficacy during aviation training. Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta.
- *Delucchi, M., & Pelowski, S. 2000. Liking or learning? The effect of instructor likeability and student perceptions of learning on overall ratings of teaching ability. *Radical Pedagogy*, 2: 1–15.
- *Dixon, N. M. 1990. The relationship between trainees' re-

- sponses on participant reaction forms and posttest scores. *Human Resource Development Quarterly*, 1: 129–137.
- *Dobransky, N. D., & Frymier, A. B. 2004. Developing teacher-student relationships through out of class communication. *Communication Quarterly*, 52: 211–223.
- *Doleys, E. J., & Renzaglia, G. A. 1963. Accuracy of student prediction of college grades. *Personnel and Guidance Journal*, 41: 528–530.
- Dunning, D., Heath, C., & Suls, J. M. 2004. Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3): 69–106.
- *Elvers, G. C., Polzella, D. J., & Graetz, K. 2003. Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology*, 30: 159–162.
- *Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. 1995. The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management*, 21: 1–25.
- Falchikov, N., & Boud, D. 1989. Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59: 395–430.
- Farh, J. L., Dobbins, G. H., & Cheng, B. S. 1991. Cultural relativity in action: a comparison of self-ratings made by Chinese and U.S. Workers. *Personnel Psychology*, 44: 129–147.
- Farh, J. L., & Werbel, J. D. 1986. Effects of purpose of appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology*, 71: 527–529.
- Feldman, K. A. 2007. Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based approach*: 93–129. Netherlands: Springer.
- Festinger, L. 1954. A theory of social comparison processes. *Human Relations*, 7: 117–140.
- Fleishman, E. A., & Quaintance, M. K. 1984. *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.
- Fleiss, J. L. 1981. *Statistical methods for rates and proportions*. New York: Wiley.
- *Fleming, K. J. 1999. Assessing course understanding in a first-person psychology course: A construct validation study of the ordered tree technique. (Doctoral dissertation, University of Notre Dame, 1999). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 59: 4460.
- *Forbes, K. J. 1988. *Building math self-efficacy: A comparison of interventions designed to increase math/statistics confidence in undergraduate students*. Unpublished doctoral dissertation, University of Florida, Gainesville.
- Franklin, B. 1750. *Poor Richard's almanac*. Philadelphia, PA: author.
- *Freedman, R. D., & Stumpf, S. A. 1978. Student evaluations of courses and faculty based on a perceived learning criterion: Scale construction, validation, and comparison of results. *Applied Psychological Measurement*, 2: 189–202.
- *Frymier, A. B. 1994. A model of immediacy in the classroom. *Communication Quarterly*, 42: 133–144.
- *Frymier, A. B., & Shulman, G. M. 1994, November. *Development and testing of the learner empowerment instrument in a communication based model*. Paper presented at the 80th Annual Meeting of the Speech Communication Association, New Orleans, LA.
- Fuhrmann, B. S., & Weissburg, M. J. 1978. Self-assessment. In M. K. Morgan & D. M. Irby (Eds.), *Evaluating clinical competence in the health professions*: 139–150. St Louis, MO: C. V. Mosby.
- *Fuqua, D. R., Johnson, A. W., Newman, J. L., Anderson, M. W., & Gade, E. M. 1984. Variability across sources of performance ratings. *Journal of Counseling Psychology*, 31: 249–252.
- *Gaier, E. L. 1961. Student self estimates of final course grades. *Journal of Genetic Psychology*, 98: 63–67.
- *Gardner, R. C., & Masgoret, A. M. 1999. Home background characteristics and second language learning. *Journal of Language and Social Psychology*, 18: 419–437.
- *Gardner, R. C., Moorcroft, R., & Metford, J. 1989. Second language learning in an immersion programme: Factors influencing acquisition and retention. *Journal of Language and Social Psychology*, 8: 287–305.
- *Gardner, R. C., Tremblay, P. F., & Masgoret, A. 1997. Towards a full model of second language learning: An empirical investigation. *Modern Language Journal*, 81: 344–362.
- *Geertshuis, S., Holmes, M., Geertshuis, H., Clancy, D., & Bristol, A. 2002. Evaluation of workplace learning. *Journal of Workplace Learning*, 14: 11–18.
- *Ghetaldi, L. R. 1998. The effect of self-modeling on climber self-efficacy, motivation, actual and perceived rock climbing skills, and knowledge in beginning rock climbers. Unpublished doctoral dissertation, University of Colorado, Greeley.
- *Greene, B. A., & Miller, R. B. 1996. Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Measurement*, 21: 181–192.
- Greenwald, A. G. 1997. Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52: 1182–1186.
- *Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. 2000. Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92: 160–170.
- *Hardre, P. L. B. 2002. The effects of instructional design professional development on teaching performance, perceived competence, self-efficacy, and effectiveness. (Doctoral dissertation, University of Iowa, 2002). *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 63: 4220.
- *Hatfield, R. D. 1996. The goal paradox in training design: Do goals operate as a narrowing or light switch mechanism? Unpublished doctoral dissertation, Indiana University, Bloomington.
- Hedges, L. V., & Olkin, I. 1985. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- *Heilenman, L. K. 1990. Self-assessment of second language ability: The role of response effects. *Language testing*, 7: 174–201.
- *Hess, J. A., & Smythe, M. J. 2001. Is teacher immediacy related to student cognitive learning? *Communication Studies*, 52: 197–219.
- Holzbach, R. L. 1978. Rater bias in performance ratings: superior, self, and peer ratings. *Journal of Applied Psychology*, 63: 579–588.

- *Hornick, S., & Tupchiy, A. 2006. Culture's impact on technology mediated learning: The role of horizontal and vertical individualism and collectivism. *Journal of Global Information Management*, 14(4): 31–56.
- Huffcutt, A. I., & Arthur, W. 1995. Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80: 327–334.
- Hunter, J. E., & Schmidt, F. L. 2004. *Methods of meta-analysis: Correcting for bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- International Training & Education Center on HIV (I-TECH). 2006, August. *Instructional evaluation categories*. Retrieved January 30, 2008 from www.go2itech.org/HTML/CM06/toolkit/monitoring/print/methods/InstrEvalMeth.doc
- *Irvine, D. W. 1965. Estimated grades and freshman achievement. *Vocational Guidance Quarterly*, 13: 193–196.
- Kanfer, R., & Kanfer, F. H. 1991. Goals and self-regulation: Applications of theory to work settings. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement*, 7: 287–326. Greenwich, CT: JAI Press.
- *Kanu, Y. 2003. Leadership development training transfer: A case study assessment of exterior post training factors of a year long leadership development program. Unpublished master's thesis, Simon Fraser University, British Columbia, Canada.
- *Keefer, K. E. 1969. Self-prediction of academic achievement by college students. *Journal of Educational Research*, 63: 53–56.
- *Keller, T. E., Whittaker, J. K., & Burke, T. K. 2001. Student debates in policy courses: Promoting policy practice skills and knowledge through active learning. *Journal of Social Work Education*, 37: 343–355.
- Kirkpatrick, D. L. 1998. *Evaluating training programs: The four levels* (2nd ed.). San Francisco, CA: Berrett-Koehler.
- *Koerner, C. D. 1991. The relationship between students' communication expectations, experiences, and subsequent perceptions of learning, interaction, and course/instructor satisfaction (Doctoral dissertation, University of Nebraska, 1991). *Dissertation Abstracts International*, 52(A): 1572.
- *Kovacs, R., & Kapel, D. E. 1976. Personality correlates of faculty and course evaluations. *Research in Higher Education*, 5: 335–344.
- Kraiger, K. 2002. Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and maintaining effective training and development: State-of-the-art lessons for practice*: 331–376. San Francisco, CA: Jossey-Bass.
- Kraiger, K., Ford, J. K., & Salas, E. 1993. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78: 311–328.
- *Krawitz, R. 2004. Borderline personality disorder: Attitudinal change following training. *Australian and New Zealand Journal of Psychiatry*, 38: 554–559.
- Kruger, J., & Dunning, D. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77: 1121–1134.
- *Kubeck, J. E., Miller-Albrecht, S. A., & Murphy, M. D. 1999. Finding information on the World Wide Web: Exploring older adults' exploration. *Educational Gerontology*, 25: 167–183.
- *La Pointe, D. K., & Gunawardena, C. N. 2004. Developing, testing and refining of a model to understand the relationship between peer interaction and learning outcomes in computer-mediated conferencing. *Distance Education*, 25: 83–106.
- *Le Rouzie, V., Ouchi, F., & Zhou, C. 1999, November. *Measuring "What People Learned" versus "What People Say They Learned": Does the Difference Matter?* Paper presented at the annual meeting of the American Evaluation Association, Orlando, FL.
- *Ledesma, E. M. 1994. Outcomes of workplace literacy training in a Midwest industrial center. Unpublished doctoral dissertation, Iowa State University, Ames.
- *Lengnick-Hall, C. A., & Sanders, M. M. 1997. Designing effective learning systems for management education: Student roles, requisite variety, and practicing what we teach. *Academy of Management Journal*, 40: 1334–1368.
- Levine, E. L., Flory, A., Ash, R. A. 1977. Self-assessment in personnel selection. *Journal of Applied Psychology*, 62: 428–435.
- Ley, K., & Young, D. B. 2001. Instructional principles for self-regulation. *Educational Technology Research and Development*, 49: 93–103.
- *Liles, A. R. 1991. The relationship between self-evaluative complexity, attributions, and perceptions of competence for achievement behavior. (Doctoral dissertation, Duke University, 1991). *Dissertation Abstracts International*, 52: 5010.
- *Lim, D. H., & Morris, M. L. 2006. Influence of trainee characteristics, instructional satisfaction, and organizational climate on perceived learning and training transfer. *Human Resource Development Quarterly*, 17: 85–115.
- Mabe, P. A., & West, S. G. 1982. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 76: 280–296.
- *MacGregor, C. J. 2000. Does personality matter: A comparison of student experiences in traditional and online classrooms. Unpublished doctoral dissertation, University of Missouri, Columbia.
- *MacIntyre, P. D., Baker, S. C., Clement, R., & Donovan, L. A. 2003. Talking in order to learn: Willingness to communicate and intensive language programs. *Canadian Modern Language Review*, 59: 589–607.
- *MacIntyre, P. D., & Charos, C. 1996. Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, 15: 3–26.
- *Maki, W. S., & Maki, R. H. 2002. Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8: 85–98.
- Marsh, H. W. 1980. The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, 17: 219–237.
- Marsh, H. W. 1983. Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75: 150–166.
- *Martens, R., & Kirschner, P. A. 2004, October. *Predicting intrinsic motivation*. Paper presented at the annual meeting of the Association for Educational Communications and Technology, Chicago, IL. (ERIC Document Reproduction Service No. ED 485061.)

- *Martineau, J. W. 1995. A contextual examination of the effectiveness of a supervisory skills training program. Unpublished doctoral dissertation, Pennsylvania State University, State College.
- *Matthews, L. M. 1997. An investigation of the use of post-training intervention to promote generalization of interpersonal skill. (Doctoral dissertation, University of Washington, 1996). *Dissertation Abstracts International*, 57(A): 5221.
- *McAlister, G. 2001. Computer-mediated immediacy: A new construct in teacher-student communication for computer-mediated distance education. Unpublished doctoral dissertation, Regent University, Virginia Beach, VA.
- McKeachie, W. J. 1997. Student ratings-the validity of use. *American Psychologist*, 52: 1218-1225.
- *McMorris, R. F., & Ambrosino, R. J. 1973. Self-report predictors: A reminder. *Journal of Education Measurement*, 10: 13-17.
- *Mihal, W. L., & Graumenz, J. L. 1984. An assessment of the accuracy of self-assessment for career decision making. *Journal of Vocational Behavior*, 25: 245-253.
- *Miles, M. B. 1965. Changes during and following laboratory training: A clinical-experimental study. *Journal of Applied Behavioral Science*, 1: 215-242.
- Miller, C. C., Glick, W. H., Wang, Y. D., & Huber, G. P. 1991. Understanding technology structure relationships: Theory development and meta-analytic theory testing. *Academy of Management Journal*, 34: 370-399.
- *Miller, R. B., Behrens, J. T., Greene, B. A., & Newman, D. 1993. Goals and perceived ability: Impact on student valuing, self-regulation, and persistence. *Contemporary Educational Psychology*, 18: 2-14.
- *Morrison, T. L., Thomas, M. D., & Weaver, S. J. 1973. Self-esteem and self-estimates of academic performance. *Journal of Consulting and Clinical Psychology*, 41: 412-415.
- *Morton, J. B., & Macbeth, W. A. A. G. 1977. Correlations between staff, peer, and self assessments of fourth-year students in surgery. *Medical Education*, 11(3): 167-170.
- *Mullen, G. E. 2003. A model of college instructors' demandingness and responsiveness and effects on students' achievement outcomes (Doctoral dissertation, Texas Tech University, 2003). *Dissertation Abstracts International*, 64(A): 1972.
- Murphy, K. R. 2008. Perspectives on the relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology*, 1: 197-205.
- *Ni, S.-F. 1994. Teacher verbal immediacy and sense of classroom community in online classes: Satisfaction, perceived learning, and online discussion. Unpublished doctoral dissertation, University of Kansas.
- Nilsen, D., & Campbell, D. P. 1993. Self-observer rating discrepancies: once an overrater always an overrater? *Human Resource Management*, 32: 265-281.
- *Noble, R. L. 2002. An assessment of student cognition in basic instruction bowling classes. Unpublished doctoral dissertation, West Virginia University, Morgantown.
- *Noels, K. A., Clement, R., & Pelletier, L. G. 1999. Perceptions of teachers' communicative style and students' intrinsic and extrinsic motivation. *Modern Language Journal*, 83: 23-34.
- *Oskarsson, M. 1981a. Self-evaluation in foreign language learning (L'auto-évaluation dans l'apprentissage des langues pour adultes: Quelques résultats de recherches). *Etudes de linguistique appliquée*, 41: 102-115.
- *Oskarsson, M. 1981b. Subjective and objective assessment of foreign language performance. In J. Read (Ed.), *Directions in language testing*. Singapore: Singapore University Press.
- *Oskarsson, M. 1984. *Self-assessment of foreign language skills: A survey of research and development work*. Paper presented to the Council for Cultural Cooperation, Strasbourg, France.
- Patiar, A., & Mia, L. 2008. The effect of subordinates' gender on the difference between self-ratings, and superiors' ratings, of subordinates' performance in hotels. *International Journal of Hospitality Management*, 27: 53-64.
- *Pena-Shaff, J., Altman, W., & Stephenson, H. 2005. Asynchronous online discussions as a tool for learning: Students' attitudes, expectations, and perceptions. *Journal of Interactive Learning Research*, 16: 409-430.
- *Peterson, S. J. 1994. *Interactive television: Continuing education participant satisfaction*. Unpublished doctoral dissertation, University of Missouri, Columbia.
- Pfeffer, J., & Sutton, R. I. 1999. *The knowing-doing gap: How smart companies turn knowledge into action*. Cambridge, MA: Harvard Business School Press.
- Pintrich, P. R. 1990. Implications of psychological research on student learning and college teaching for teacher education. In R. Houston (Ed.), *The handbook of research on teacher education*: 826-857. New York: Macmillan.
- *Poorman, S. G. 1988. Levels of test anxiety and cognitions of second semester senior level baccalaureate nursing students preparing for licensure exam. Unpublished doctoral dissertation, University of Pittsburgh, PA.
- *Poteet, M. L. 1996. The training transfer process: An examination of the role of individual, motivational, and work environmental factors. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.
- *Pribyl, C. B., Sakamoto, M., & Keaten, J. A. 2004. The relationship between nonverbal immediacy, student motivation, and perceived cognitive learning among Japanese college students. *Japanese Psychological Research*, 46: 73-85.
- *Quinones, M. 1995. Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology*, 80: 226-238.
- *Radhakrishnan, P., Arrow, H., & Snizek, J. A. 1996. Hoping, performing, learning, and predicting: Changes in the accuracy of self-evaluations of performance. *Human Performance*, 9: 23-49.
- *Rambo, G. A. 1997. Relationships among self-efficacy, anxiety, perceptions of clinical instructor effectiveness and senior baccalaureate nursing students' perceptions of learning in the clinical environment. Unpublished doctoral dissertation, The University of Oklahoma, Norman.
- *Ramirez, A. E. 2000. Individual, attitudinal, and organizational influences on training effectiveness: A test of Noe's model. Unpublished doctoral dissertation, University of Tulsa, OK.
- *Raven, B. H., & Fishbein, M. 1965. Social referents and self-evaluation in examinations. *Journal of Social Psychology*, 65: 89-99.
- *Richardson, J. C., & Swan, K. 2003. Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of the Asynchronous Learning Network*, 7: 68-88.

- *Richmond, V. P. 1990. Communication in the classroom: Power and motivation. *Communication Education*, 39: 181–195.
- *Robinson, R. Y. 1995, May. *Affiliative communication behaviors: A comparative analysis of the interrelationships among teacher nonverbal immediacy, responsiveness, and verbal receptivity on the prediction of student learning*. Paper presented at the International Communication Association Annual Meeting, Albuquerque, NM.
- *Rodriguez, J. I., Plax, T. G., & Kearney, P. 1996. Clarifying the relationship between teacher nonverbal immediacy and student cognitive learning: Affective learning as the central causal mediator. *Communication Education*, 45: 293–305.
- *Roth, B. F. 1996. Student outcomes assessment of Air Command and Staff College: An evaluative study. Unpublished doctoral dissertation, University of Virginia, Charlottesville.
- *Rovai, A. P., & Baker, J. D. 2005. Gender differences in online learning: Sense of community, perceived learning, and interpersonal interactions. *The Quarterly Review of Distance Education*, 6: 31–44.
- *Rubin, R. B., Graham, E. E., & Mignerey, J. T. 1990. A longitudinal study of college students' communication competence. *Communication Education*, 39: 1–14.
- Rynes, S. L., Colbert, A. E., & Brown, K. G. 2002. HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41: 149–174.
- *Saleh, U. A. 1994. Motivational orientation and understanding of limits and continuity of first year calculus students. Unpublished doctoral dissertation, The University of Oklahoma, Norman.
- *Schneider, P. H. 2001. Pair taping: Increasing motivation and achievement with a fluency practice. *Teaching English as a Second or Foreign Language*, 5. Retrieved November 9, 2007 from <http://www-writing.berkeley.edu/tesl-ej/ej18/a2.html>
- *Schunk, D. H., & Ertmer, P. A. 1999. Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Journal of Educational Psychology*, 91: 251–260.
- *Sekowski, G. J. 2002. Evaluating training outcomes: Testing an expanded model of training outcome criteria. Unpublished doctoral dissertation, DePaul University, Chicago, IL.
- *Senko, C. M. 2003. The stability and regulation of achievement goals: The role of competence feedback. (Doctoral dissertation, University of Wisconsin, Madison, 2002). *Dissertation Abstracts International*, 63: 5570.
- *Shin, N. 2003. Transactional presence as a critical predictor of success in distance learning. *Distance Education*, 24: 69–86.
- *Sim, J. M. 1990. Relationships among self-perception of competency, informational & performance competencies, and self-actualization. Unpublished doctoral dissertation, University of San Francisco, CA.
- *Sinkavich, F. J. 1995. Performance and metamemory: Do students know what they don't know? *Journal of Instructional Psychology*, 22: 77–87.
- *Sitzmann, T. 2005a. Improving learning outcomes in learner controlled training environments: Will prompting self regulation and delaying trainees' judgments of learning make a difference? Unpublished doctoral dissertation, University of Tulsa, OK.
- *Sitzmann, T. 2005b. [NATO school pilot evaluation data.] Unpublished raw data.
- *Sitzmann, T. 2006. *Final report on NATO school evaluation*. Alexandria, VA: Advanced Distributed Learning Co-Laboratory.
- *Sitzmann, T. 2007. [Web-based training evaluation data.] Unpublished raw data.
- Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. 2008. A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93: 280–295.
- *Sitzmann, T., Brown, K. G., Ely, K., & Kraiger, K. 2009. A cyclical model of motivational constructs in Web-based courses. *Military Psychology*, 21: 534–551.
- Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. 2006. The comparative effectiveness of Web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, 59: 623–664.
- *Sniezek, J. A. 1998. Charting the course of self-evaluations and social comparisons over time. Unpublished manuscript.
- *Spiros, R. K. 2003. Individual differences in motivation during distance training: The influence of goal orientation and self-efficacy on learning outcomes (Doctoral dissertation, University of California, 2003). *Dissertation Abstracts International*, 63(B): 6131.
- *Stark, R., Gruber, H., Renkl, A., & Mandl, H. 1998. Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction*, 8: 117–129.
- *Stefani, L. A. J. 1994. Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19: 69–75.
- Stewart, G. L. 1999. Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology*, 84: 959–968.
- Stuart, M. R., Goldstein, H. S., & Snope, F. C. 1980. Self-evaluation by residents in family medicine. *Journal of Family Practice*, 10: 639–642.
- *Stumpf, S. A., & Freedman, R. D. 1979. Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, 71: 293–302.
- *Sullivan, M. E., Hitchcock, M. A., & Dunnington, G. L. 1999. Peer and self assessment during problem based tutorials. *American Journal of Surgery*, 177: 266–269.
- *Sullivan, T. A., Sharma, M., & Stacy, R. 2002. Effects of a brief training program for lay health volunteers to facilitate smoking cessation among African Americans. *Journal of Alcohol and Drug Education*, 47: 4–17.
- *Svanum, S., & Bigatti, S. 2006. Grade expectations: Informed or uninformed optimism, or both? *Teaching of Psychology*, 33: 14–18.
- *Swan, K. 2001. Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. *Distance Education*, 22: 306–331.
- *Tan, J. A., Hall, R. J., & Boyce, C. A. 2003. The role of employee reactions in predicting training effectiveness. *Human Resource Development Quarterly*, 14: 397–411.
- *Teven, J. J. 1998. The relationships among teacher characteristics, student learning, and teacher evaluation. Unpublished doctoral dissertation, West Virginia University, Morgantown.
- *Teven, J. J. 2005. Teacher socio-communicator style and toler-

- ance for disagreement and their association with student learning in the college classroom. *Texas Speech Communication Journal*, 30(1): 23–35.
- *Teven, J. J., & McCroskey, J. C. 1996. The relationship of perceived teacher caring with student learning and teacher evaluation. *Communication Education*, 46: 1–9.
- *Thompson, C. A. C. 1992. The relationships among teachers' immediacy behaviors, credibility, and social style and students' motivation and learning: Comparisons among cultures. Unpublished doctoral dissertation, West Virginia University, Morgantown.
- *Torok, S. E., McMorris, R. F., & Lin, W. C. 2004. Is humor an appreciated teaching tool? Perceptions of professors' teaching styles and use of humor. *College Teaching*, 52: 14–20.
- *Tousignant, M., & DesMarchais, J. E. 2002. Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: A correlation study. *Advances in Health Sciences Education*, 7: 19–27.
- Tracey, J. M., Arroll, B., Barham, P. M., & Richmond, D. E. 1997. The validity of general practitioners' self assessment of knowledge: Cross sectional study. *BMJ*, 315: 1426–1428.
- *Tseng, M. S. 1972. Self-perception and employability: A vocational rehabilitation problem. *Journal of Counseling Psychology*, 19: 314–317.
- *Tziner, A., Haccoun, R. R., & Kadish, A. 1991. Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. *Journal of Occupational Psychology*, 64: 167–177.
- *Van Cleave, A. K., Jr. 1995. Engagement and training outcomes: Extending the concept of motivation to learn. Unpublished doctoral dissertation, University of Tennessee, Knoxville.
- *Wade, G. H. 2001. Perceptions of instructor caring behaviors, self-esteem, and perceived clinical competence: A model of the attitudinal component of professional nurse autonomy in female baccalaureate nursing students. Unpublished doctoral dissertation, Widener University School of Nursing, Chester, PA.
- *Walczyk, J. J., & Hall, V. C. 1989. Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology*, 81: 435–437.
- *Walker, K. B., & Hackman, M. Z. 1991, October. *Information transfer and nonverbal immediacy as primary predictors of learning and satisfaction in the televised course*. Paper presented at the annual meeting of the Speech Communication Association, Atlanta, GA.
- *Wangsotorn, A. 1980. Self-assessment in English skills by undergraduate and graduate students in Thai Universities. In J. Read (Ed.), *Directions in language testing*: 240–260. Singapore: Singapore University Press. (Also titled Self-assessment of language proficiency [mimeo]. Singapore: SEAMEO/REL. [1982]).
- *Wanzer, M. B., & Frymier, A. B. 1999. The relationship between student perceptions of instructor humor and student's reports of learning. *Communication Education*, 48: 48–62.
- *Warr, P., Allan, C., & Birdi, K. 1999. Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72: 351–375.
- Weiss, H. M., & Cropanzano, R. 1996. Affective Events Theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior*, 18: 1–74.
- *Wesche, M. B., Morrison, F., Ready, D., & Pawley, C. 1990. French immersion: Postsecondary consequence for individuals of universities. *Canadian Modern Language Review*, 46: 430–451.
- *Wexley, K. N., & Baldwin, T. T. 1986. Post-training strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal*, 29: 503–520.
- *Wheeler, A. E., & Knoop, H. R. 1982. Self, teacher, and faculty assessments of student teaching performance. *Journal of Educational Research*, 75: 171–181.
- Whitener, E. M. 1990. Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75: 315–321.
- *Wilkerson, L., Lee, M., & Hodgson, C. S. 2002. Evaluating curricular effects on medical students' knowledge and self-perceived skills in cancer prevention. *Academic Medicine*, 77: 51–53.
- *Williams, G. C., & Deci, E. L. 1996. Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, 70: 767–779.
- *Wilson, R. H. 2006. The effect of one leadership training program on behavioral and skill change in community college division chairs and other organizational leaders. Unpublished doctoral dissertation, University of Nebraska, Lincoln.
- Winne, P. H. 1995. Inherent details in self-regulated learning. *Educational Psychologist*, 30: 173–187.
- *Wise, A., Chang, J., Duffy, T., & Del Valle, R. 2004. The effects of teacher social presence on student satisfaction, engagement, and learning. *Journal of Educational Computing Research*, 31: 247–271.
- *Witt, P. L., & Wheelless, L. R. 2001. An experimental study of teachers' verbal and nonverbal immediacy and students' affective and cognitive learning. *Communication Education*, 50: 327–342.
- *Woolliscroft, J. O., TenHaken, J., Smith, J., & Calhoun, J. G. 1993. Medical students' clinical self-assessments: Comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Academic Medicine*, 68: 285–294.
- Yik, M. S. M., Bond, M. H., & Paulhus, D. L. 1998. Do Chinese self-enhance or self-efface: It's a matter of domain. *Personality and Social Psychology Bulletin*, 24: 399–412.
- *Young, F. C. 1954. College freshman judge their own scholastic promise. *Personnel and Guidance Journal*, 32: 399–403.
- *Zhao, H., Seibert, S. E., & Hills, G. E. 2005. The mediating role of self-efficacy in the development of entrepreneurial intentions. *Journal of Applied Psychology*, 90: 1265–1272.
- Zimmerman, B. J. 1990. Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25: 3–17.
- Zimmerman, R. D., Mount, M. K., & Goff, M., III. 2008. Multisource feedback and leaders' goal performance: Moderating effects of rating purpose, rater perspective, and performance dimension. *International Journal of Selection and Assessment*, 16, 121–133.

Traci Sitzmann is an assistant professor of management at the University of Colorado, Denver. She completed her PhD at the University of Tulsa and subsequently provided consulting services to the U.S. Department of Defense for improving the quality of online training. Sitzmann's research focuses on self-regulated learning and attrition from training.

Katherine Ely is a senior research associate at Fors Marsh Group in Arlington, Virginia. She earned her PhD in industrial and organizational psychology at George Mason University. Her research focuses on organizational training and development, including leadership development and training designs that promote adaptability.

Kenneth G. Brown is Henry B. Tippie Research Fellow and associate professor of management and organizations at the University of Iowa. He received his PhD from Michigan State University and is certified as a senior professional in human resource management. Brown's research focuses on motivation, training effectiveness, and learning.

Kristina N. Bauer is a doctoral candidate at Old Dominion University. She completed her master's in human resource management at George Washington University. Her research interests include self-regulation, individual differences, and technology-delivered instruction.

Copyright of Academy of Management Learning & Education is the property of Academy of Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.