Self-Assessments of Knowledge: Where do we go from Here?

Bradford S. Bell

Jessica E. Federman

Cornell University

Citation:

Bell, B. S., & Federman, J. E. (2010). Self-assessments of knowledge: Where do we go from here? *Academy of Management Learning & Education, 9 (2),* 342-347.

Self-Assessments of Knowledge: Where do we go from Here?

Sitzmann, Ely, Brown, and Bauer (this issue) provide a much needed review and integration of the literature on self-assessments of knowledge. They use meta-analytic methods to sort through a body of work characterized by mixed findings and competing conclusions. Their findings provide convincing evidence that trainees generally have difficulty self-assessing their learning. Based on these findings, Sitzmann et al. recommend a more limited role for self-assessments in evaluation research and practice, but stop short of suggesting that self-assessments should be abandoned as a tool to measure learning.

In this paper, we argue that there remain several unanswered questions surrounding self-assessments of knowledge that must be addressed before we can reach a more definitive conclusion on the viability of these measures. The answers to these questions may provide further evidence that self-assessments should not be used as an indicator of learning or they may serve to qualify the conditions under which self-assessments can be used with reasonable confidence. In either case, addressing these issues is critical if work in this area is to influence how researchers and practitioners evaluate trainees' learning.

Different Perspectives on Self-Assessments

In a recent article published in *T+D*, one of the leading publications for training practitioners, Lopker and Askeland (2009) discuss how organizations can effectively use Level 1 evaluation (i.e., measures of participants' reactions) to assess their training. However, they also note that Level 1 evaluation should not be used alone and recommend incorporating components of Level 2 evaluation (i.e., measures of learning) to strengthen the assessment. They highlight several different methods that can be used to assess learning, but argue that "One of the easiest is to have learners self-report their level of knowledge at the beginning of the training and then

again at the end of training" (p. 75). Another option they recommend is "At the end of the training, ask learners to quantify their learning improvement …" (p. 75).

The findings and recommendations presented by Sitzmann et al. (this issue) stand in stark contrast to those offered by Lopker and Askeland (2009). Sitzmann and colleagues' meta-analysis revealed only a moderate correlation between learners' self-assessments of their knowledge level and measures of cognitive learning ($\rho = .44$) and no correlation ($\rho = .00$) between learners' self-assessments of their knowledge gain (i.e., improvement) and cognitive learning. Raising further concerns about the construct validity of these self-report measures, they also found that self-assessments of knowledge correlated more strongly with affective than cognitive training outcomes and that a majority of studies reached the conclusion that the self-assessments were inaccurate. Based on these findings, Sitzmann et al. recommend that researchers and practitioners "… be more prudent in their use of self-assessments of knowledge as a cognitive learning measure" (p. 30) and rely on objective tests, rather than self-report measures, to assess learners' knowledge level and gain.

At first glance it is surprising that two articles could offer such different recommendations on the use of self-assessments of knowledge. One may be tempted to conclude that this is yet another example of the science-practice gap that exists in the HR field in general (Rynes, Colbert, & Brown, 2002) and the field of training evaluation more specifically (Kraiger, 2002). However, a closer examination of the literature on self-assessments reveals that this body of work is characterized by mixed findings and conflicting conclusions, suggesting that the problem is not a science-practice gap but rather a gap in the science itself. For instance, Sitzmann et al. examined the conclusions that studies reached regarding the accuracy of self-assessment of knowledge and found that 56% of studies concluded that self-assessments are

inaccurate. However, they also found that 26% of studies concluded self-assessments were accurate and 18% reported mixed results. Given that the literature has failed to reach consensus regarding the accuracy of self-assessments, it is not surprising that Sitzmann et al. also find that many researchers continue to interpret self-assessments of knowledge as a valid indicator of learning. Nor is it surprising that these mixed results appear to have had little effect on practitioners' use of self-assessments as a tool to measure learning.

The study by Sitzmann et al. provides some resolution to these discrepant research findings and helps to clarify the practical implications of using self-assessments to measure knowledge. In contrast to the mixed conclusions reached by the individual studies, their meta-analytic findings paint a consistent picture and suggest that self-assessments of knowledge are generally not a valid measure of cognitive learning and should not be used to assess learners' knowledge level or improvement. Yet, their review also exposes several gaps in our understanding of self-assessments of knowledge. In particular, questions persist concerning the validity of different types of self-assessments of knowledge, the contextual conditions that influence the validity of these self-assessments, and the extent to which this research generalizes to applied settings. We believe these unresolved issues underlie not only the conflicting findings reported by studies in this area but also the failure of this research to drive change in how practitioners assess learners' knowledge. In the following section, we discuss each of these issues and their implications for future research in this area.

Teasing Apart Different Types of Self-Assessments of Knowledge

Sitzmann et al. define self-assessments of knowledge as "… the evaluations learners make about their current knowledge levels or increases in their knowledge levels in a particular domain." This definition is consistent with how self-assessments are typically used to measure

learning in organizational settings. Learners either self-report their knowledge pre- and post-training or are asked to assess their knowledge gain at the end of training. However, as Sitzmann et al. point out, past research in this area has examined a diversity of self-assessments. Based on a review of the articles included in the meta-analysis, we have identified at least three important differences in these self-assessments, which center on issues of construct dimensionality, context scope, and temporal focus. For example, the specific constructs that learners are asked to evaluate varies across the studies. In some studies, learners assess their perceived learning or comprehension (e.g., Le Rouzie, Ouchi, and Zhou, 1999; Maki & Maki, 2002), whereas in others they evaluate their performance (e.g., Chur-Hansen, 2000; Quiñones, 1995) or competence (e.g., Arnold, 1985; Carrell & Willmington, 1996). The scope of the self-assessments also varies, such that some studies ask learners to judge their perceived learning or performance in a single course (e.g., Baird, 1987) and others ask learners to assess their perceived learning across multiple courses or an entire curriculum (e.g., Chemers, Hu, & Garcia, 2001; Zhao, Seibert, & Hills, 2005). Furthermore, there exist differences in the temporal focus of the self-assessments examined across studies. Some studies ask learners to evaluate their past performance (e.g., Quiñones, 1995) and others ask learners to assess their future or expected performance (Doleys & Renzaglia, 1963).

Sitzmann et al. distinguish between assessments that focus on absolute knowledge level and those that focus on knowledge gain and find that the former are generally a more valid indicator of cognitive learning. An important question, however, is whether there are other meaningful differences across these measures that influence their validity. Are there differences in learners' ability to assess their perceived learning versus their perceived competence, their learning in a single course as opposed to across multiple courses, or their past performance

relative to their future performance? The issue is not simply the extent to which a given assessment is similar to a corresponding outcome measure, although that is an important consideration. It is also important to consider whether certain types of measures yield more accurate self-assessments of knowledge and thus serve as better predictors of cognitive learning than others. Carpenter, Friar, and Lipe (1993), for instance, examined expectations of academic success among students in introductory accounting courses. In the third week of the course, they asked students to report their expected grade. They found that minority students had lower grade expectations than majority students, but the correlation between expected and actual course grades was significantly stronger for majority students than minority students. Similarly, they found that female students had lower grade expectations than male students, but the two groups did not significantly differ in their actual course grades. Based on these findings, the authors conclude that minority and female students' course performance (a measure of cognitive learning) does not align well with their grade expectations, in part because these expectations are influenced by factors such as the students' past scholastic performance. The question is whether the same pattern of results would have emerged if the self-assessment of knowledge had asked students to rate their perceived learning relative to specific course objectives; a judgment that may be less susceptible to some of the factors thought to influence minority students' grade expectations and may, as a result, represent a more accurate self-assessment of their knowledge and better predictor of their cognitive learning.

In another study, Chemers et al. (2001) examined the effects of academic self-efficacy and optimism on first year college students' academic performance, stress, health, and commitment to remain in school. During the first week of the winter quarter, before receiving a detailed evaluation of their performance, students were asked to self-assess their recent academic

performance on a 5-point scale. Compared to studies that have asked students to assess their learning in a single course (Baird, 1987) or lecture (e.g., Chesboro & McCroskey, 2000), the self-assessment examined in the Chemers et al. study is much broader in its focus (overall performance during the first academic quarter). In addition, whereas students in the Carpenter et al. (1993) study described above reported their grade expectations early in the course, students in the Chemers et al. study made a retrospective judgment of their academic performance after their courses had ended. Although there are clear differences in the scope and temporal focus of the self-assessments examined across these studies, the question is the extent to which these differences influence the validity of learners' judgments of their knowledge.

Ultimately, to answer these questions future research is needed that teases apart differences across these various types of self-assessments in terms of their construct dimensionality, context scope, and temporal focus, and examines the extent to which these differences influence the accuracy of learners' self-assessments of knowledge. A good model for the type of research we are suggesting is a study by Radhakrishnan, Arrow, and Sniezek (1996), in which the authors examined the extent to which the accuracy of students' self-evaluations of their exam performance varied as a function of the temporal context (pre-performance vs. post-performance evaluation) and temporal distance (time between the judgment and performance). Unfortunately, most studies in this area have focused on a single type of self-assessment, which controls away the very differences that we argue may influence the validity of these measures. Research that examines these three issues may not only help to address the questions raised above but may also have important theoretical and practical implications. Theoretically, this research may help identify the different mechanisms that explain why individuals struggle to accurately self-assess their knowledge. For example, students' self-concept or optimism may

skew their academic expectations (Carpenter et al., 1993; Chemers et al., 2001), whereas a lack of meta-cognitive skills may bias their perceived competence (Kruger & Dunning, 1999). At a more fundamental level, this research may help to resolve the varying conceptual definitions that exist within the self-assessment literature, none of which makes explicit the three factors we have discussed. Practically, this research may help training professionals design more valid self-assessments of knowledge or develop interventions aimed at enhancing the accuracy of learners' assessments. Given the current diversity of self-assessments that have been examined and lingering questions concerning their differential validity, it is difficult to provide practitioners with firm recommendations around how, if at all, they should be using self-assessments of knowledge.

Further Clarifying the Role of the Context

In addition to examining the accuracy of different types of self-assessments, it is important to understand the conditions under which learners' self-assessments are more or less valid. That is, under what conditions are learners able to accurately judge their knowledge? This question has been relatively neglected within the literature, however, as most studies have focused on the more fundamental question of whether or not learners' self-assessments are accurate (Radhakrishnan et al., 1996). Yet, contextual differences across these studies, which are rarely examined and often not clearly articulated, may help to explain and resolve the variability in prior research findings. In addition, a better understanding of these conditions has important theoretical implications given the important role of self-evaluations in the process of self-regulation. From a practical perspective, training practitioners need guidance on whether there are situations where self-assessments can be used with reasonable confidence to evaluate

trainees' learning and they need access to tools and strategies that can be used to enhance the accuracy of learners' self-evaluations.

The study by Sitzmann et al. provides valuable insight into the role of the context in shaping the relationship between self-assessments of knowledge and measures of cognitive learning. For example, they find that this relationship is stronger in courses that provide external feedback, in classroom instruction and blended learning than in web-based instruction, and when learners practiced self-assessing their knowledge and received feedback on their accuracy. These findings provide evidence that the context matters, and suggest that future research needs further delineate the conditions that influence the accuracy of learners' self-assessments. For example, research should examine the effects of different types of feedback (e.g., descriptive, normative, velocity) as well as the timing and frequency of feedback on the accuracy of learners' self-assessments (Kozlowski et al., 2001). Also, research is needed to understand why learners' self-assessments tend to be less accurate in web-based instruction and to identify instructional features (e.g., interactivity, immersion) that can be delivered through technology to enhance these self-evaluations (Kozlowski & Bell, 2007).

Although Sitzmann et al. highlight a set of conditions that significantly enhance the accuracy of learners' self-assessments, their results also suggest that there is an opportunity to further strengthen the self-assessment-cognitive learning relationship. Specifically, the highest correlation reported between self-assessments and cognitive learning in their meta-analysis was .55, which was under an ideal set of conditions, and the moderators they examined did not fully account for the variability in the self-assessment-cognitive learning relationship. These findings suggest that future research not only needs to examine other contextual factors but also should explore specific instructional supports that can be used to further enhance the accuracy of

learners' self-assessments. For example, adaptive guidance is designed to help trainees interpret their feedback and calibrate their progress toward task mastery, which should enhance the accuracy of learners' self-evaluations (Bell & Kozlowski, 2002). Also, recent research suggests that prompting trainees to regularly reflect on their learning and comprehension may be an effective strategy for enhancing the accuracy of learners' self-evaluations (Schmidt & Ford, 2003). As training increasingly moves out of the classroom and into technology and the workplace, trainees are being given greater control over their learning. Yet, research has found that learners often do not make good use of this control – they have difficulty assessing their learning, make poor learning decisions, and withdraw from training prematurely (Brown, 2001). Thus, strategies that support learners' self-evaluation activities may not only strengthen the self-assessment-cognitive learning relationship but also help trainees make better learning choices in learner-controlled environments.

Graduating from College

A final issue that we believe warrants attention in the literature on self-assessments involves the need to extend this research to applied settings. The vast majority of the studies included in Sitzmann et al.'s meta-analysis were conducted in college settings, typically in the context of undergraduate courses. Participants in these studies are generally young adults and the courses are overwhelmingly classroom-based (rather than online) and focused on cognitive, rather than skill-based, content. Although we expect that the core findings of this research will generalize beyond the college classroom, we also believe there are several reasons to study learners' self-assessments in organizational settings.

First, as noted above, organizations are increasingly moving from classroom-based training to more online and on-the-job training. In these environments, trainees are not only

given greater control over their learning but they also may not have access to the instructional supports (e.g., external feedback) commonly found in classroom settings. Thus, it is important for future research in this area to focus more attention on understanding how different features of technology-delivered instruction and experiential training programs enhance or inhibit the accuracy of learners' self-assessments. Second, much of the training that is conducted in organizations is focused on building skills in areas such as sales, management, and customer service. In fact, the recent economic downturn has led many organizations to focus their training efforts on improving skills that are highly specific to learners' jobs and can immediately be put into practice ("Gauges & Drivers", 2008). Given that Sitzmann et al. find that course content moderates the relationship between self-assessments and cognitive learning, it will be important for future research to focus more attention on studying learners' self-assessments in more applied, skill-based training programs. Finally, recent research suggests that age-related changes in adult development may have important implications for individuals' work behavior and job performance (Kanfer & Ackerman, 2004). For example, individuals experience a decline in fluid intellectual abilities (e.g., working memory, attention) as they age. Training places significant demands on these abilities, which means that older workers may find it more difficult than younger workers to learn new skills. In addition, individuals tend to be sensitive to changes in their abilities with increasing age and typically seek to protect their self-concept. For example, older workers may avoid certain types of career development activities. We believe future research should explore the extent to which these age-related changes influence the accuracy of learners' self-assessment of knowledge. For example, older trainees may find it more difficult to devote the attentional resources necessary to monitor their learning and may also manipulate

their self-assessments, or avoid self-assessing their knowledge, in an effort to protect their self-concept.

We also believe there is value in conducting basic descriptive research designed to better understand the use of self-assessments in applied settings. Although Sitzmann et al. examine how the use and interpretation of self-assessments has changed over the past three decades, their analysis is limited by the fact that most studies have been conducted by researchers in university settings, not practitioners in applied settings. There are reasons to expect that the evaluation practices of these two groups may differ. For example, one of the most commonly cited obstacles to evaluation in organizational settings is the fact that many practitioners lack training in evaluation methods (Kraiger, 2002). The result is that practitioners may be more likely to rely on subjective assessments of learning than researchers, who typically have at least some training in areas such as evaluation design and test construction. Descriptive research that focuses on learning evaluation practices in applied settings will provide a better understanding of the science-practice gap in this area.

Conclusion

In closing, we believe the study by Sitzmann and colleagues makes a valuable contribution to our understanding of self-assessments of knowledge. It begins to resolve the discrepant findings of past research and provides some clarity around the practical implications of using self-assessments to measure trainees' cognitive learning. Perhaps more importantly, by reviewing and integrating this diverse body of work, their study reveals several gaps in our understanding of self-assessments and highlights fruitful avenues for future research in this area. We believe that future research that focuses on teasing apart different types of self-assessments, understanding the conditions under which these assessments are more and less valid, and

examines these questions in not only university settings but also work contexts can help address these gaps and advance our understanding of learners' self-assessments. Our hope is that the end result will be greater alignment of science and practice in the area of learning evaluation.

References

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.

Baird, J. S. (1987). Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology, 79,* 90-91.

Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology, 55,* 267-306.

Brown, K. G. (2001). Using computers to deliver training: Which employees learn and why? *Personnel Psychology, 54,* 271-296.

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.

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.

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.

Doleys, E. J., & Renzaglia, G. A. (1963). Accuracy of student prediction of college grades. *Personnel and Guidance Journal, 41,* 528-530.

Gauges & drivers. (2008, November/December). *Training Magazine, 45 (9),* 16-34.

Kanfer, R., & Ackerman, P. L. (2004). Aging, adult development, and work motivation. *Academy of Management Review, 29,* 440-458.

Kozlowski, S. W. J., & Bell, B. S. (2007). A theory-based approach for the designing distributed learning systems. In S. M. Fiore & E. Salas (Eds.), *Where is the learning in distance learning? Toward a science of distributed learning and training* (pp. 15-39). Washington, DC: APA Books.

Kozlowski, S. W. J., Toney, R. J., Mullins, M. E., Weissbein, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice.* San Francisco, CA: Jossey-Bass.

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.

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 atthe annual meeting of the American Evaluation Association, Orlando, FL.

Lopker, G., & Askeland, R. (2009). More than a smile sheet: Using Level 1 evaluation effectively. *T+D, 63 (9),* 74-75.

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. J*ournal of Experimental Psychology: Applied, 8,* 85-98

Quiñones, M. (1995). Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology, 80,* 226-238.

Radhakrishnan, P., Arrow, H., & Sniezek, J. A. (1996). Hoping, performing, learning, and predicting: Changes in the accuracy of self-evaluations of performance. *Human Performance, 9,* 23-49.

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.

Schmidt, A. M., & Ford, J. K. (2003). Learning within a learner control training environment: The interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology, 56,* 405-429.

Sitzmann, T., Ely, K., Brown, K. G., Bauer, K. (in press). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning and Education.*

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.