Measuring Self-Regulation in Computer-Based Learning Environments

Gregory Schraw

Department of Educational Psychology
University of Nevada, Las Vegas

I provide a summary of the four invited articles in this special issue and compare and contrast different methods for measuring self-regulation in computer-based learning environments (CBLEs). I present a taxonomy that distinguishes between offline and online measures and further distinguishes subcategories within each of these categories. I discuss four measurement challenges, including orchestrating a repertoire of automated assessments to capture these ongoing activities, developing scoring rubrics and scoring criteria, interpreting assessments, and aligning multiple outcome measures. I also discuss how measures of self-regulated learning in CBLEs can be used to assess important cognitive processes such as planning, strategy use, monitoring, and increased learning. I conclude with several suggestions for measurement practices that enhance and further strengthen the validity of research conclusions.

Self-regulated learning (SRL) refers to monitoring and controlling one’s own cognitive performance before, during, and after a learning episode. Self-regulation includes elements of planning, goal setting, strategy implementation, summarizing, and monitoring one’s progress (Azevedo, 2005; Winne & Nesbit, 2009; Zimmerman, 2008). Several different models of self-regulation that have common and distinctive features have been proposed over the two decades (Butler & Winne, 1995; Pintrich, 2000; Winne, 2001; Winne & Perry, 2000; Zimmerman, 2000, 2001). Schraw, Kauffman, and Lehman (2002) provided a review of self-regulation models and partitioned the construct into three components they referred to as knowledge, metacognition, and motivation. Knowledge included facts, concepts, and schemata related to a particular learning task. Metacognition refers to knowledge and regulation of cognition (Schraw, 2006; White & Frederiksen, 2005). Knowledge of cognition consists of knowledge about oneself as a learner as well as the conditions that constrain learning. Regulation of cognition includes a wide variety of domain-general skills such as goal setting, planning, implementing strategies, monitoring, and evaluating one’s learning. Motivation refers to beliefs such as epistemological beliefs and self-efficacy that affect engagement and persistence.

Many students are not sufficiently self-regulated (Pressley & Harris, 2006; Schunk & Zimmerman, 2006), and even good learners experience trouble regulating learning in unfamiliar domains or challenging circumstances. One special type of self-regulatory challenge occurs when learning in a hypermedia environment (Azevedo, 2005; Winne & Hadwin, 2008) because the use of hypermedia greatly increases task demands and requires the learner to stretch limited processing resources across two major constraints: to-be-learned information and the hypermedia environment. A second challenge is that hypermedia environments may be more likely to present complex, unfamiliar information that taxes learners even without additional hypermedia demands (Reed, 2006). Nevertheless, individuals of all ages are asked to use hypermedia inside and outside the classroom; thus, it is essential to develop a better conceptual understanding of the design and implementation of hypermedia to enhance learning and self-regulation (Azevedo, 2005). Yet despite added processing challenges in some situations, the benefits of CBLEs often create expanded learning opportunities for students.

Hypermedia learning environments impose special demands on researchers as well as learners. One is the design and implementation of new media intended to increase complex, integrated understanding of challenging information (Azevedo, 2007, 2008; Graesser, McNamara, & VanLehn, 2005; Greene & Azevedo, 2007, 2009; Winne & Hadwin, 2008). The present contributions discuss these challenges in detail. A second challenge is measuring the success of
such programs before, during, and after learning. Fortunately, great strides have been made in both of these areas over the past two decades. A growing body of research articles, books, and special issues of journals has addressed the design and implementation aspect of computer-based learning environments (CBLEs). However, the current issue is the first to focus exclusively on the challenges of conceptualizing and implementing measurement of online self-regulation in CBLEs.

This article is divided into five sections. This section provides a rationale for the special issue and overviews the current article. The second section summarizes each of the four invited articles and highlights their suggestions for better measurement of SRL. The third section provides a taxonomy of measurement strategies based on these articles that distinguishes between offline and online measures of self-regulation, and considers several subcategories within each. The fourth section discusses four generic challenges in the measurement of SRL, and the fifth section considers implications for improving current measures and developing new measures.

**SUMMARY OF THE ARTICLES**

Each of the four invited articles in this issue explores a different facet of the measurement of SRL. Azevedo, Moos, Johnson, and Chauncey (2010/this issue) focus on cognitive and metacognitive processes, especially the underlying control and monitoring components of metacognition. They argue that even internal cognitive and metacognitive processes can be detected, traced, modeled, and fostered during learning, using unobtrusive online trace methodologies such as eye tracking and timed evidence of error detection as well as more obtrusive measures such as think alouds and verbal reports of processing to provide convergent evidence of self-regulation (Alexander, 2003; Anderson & Lebiere, 1998; Azevedo & Wutherspoon, 2009; Ericsson, 2006). They discuss a variety of offline and online indicators of self-regulation, where offline measures refer to events that occur before or after ongoing processing, whereas online measures refer to activities that occur during processing. For present purposes, I use the term **online indicator** to refer to a measurement event that competes simultaneously with an information-processing activity. For example, thinking aloud while reading or conducting a search constitutes an online measure. I use the term **obtrusive measure** to refer to a measurement event that potentially may interfere with information processing because it competes for limited resources. Online measurements may be obtrusive in some situations but not others, for example, when assessing the information processing of novices, but not experts (Hertzum, Hansen, & Andersen, 2009; Kirk & Ashcraft, 2001).

Azevedo et al. (2010/this issue) focus on **what** processes are used to self-regulate and **why** they are used. Regarding what processes are used, some traces of cognitive activity are easily measured unobtrusively using time stamps or navigation markers as learners maneuver through a text. Other traces are more difficult to measure without engaging the learner; therefore, it may be necessary to ask them to stop and explain, keep a concurrent log, or interact with pedagogical agents. These measures provide a wealth of information and some types of information that would be impossible to gather otherwise, even though these data-collection methods may potentially change online processing for some learners. Indeed, self-report measures may provide the best window into learners’ minds in order to understand why they self-regulate as they do.

Overall, they conclude that no single measure is sufficient for measuring the rich self-regulatory behaviors and strategies used by savvy learners, in part, because self-regulation is a highly dynamic process that changes over time. They suggest a variety of measurement strategies that span all phases of self-regulation as learners prepare for learning, construct meaning, monitor, and finally integrate what they have learned. These strategies are discussed in more detail next after a review of others discussed among the four invited articles.

Aleven, Roll, McLaren, and Koedinger (2010/this issue) focus on studying ongoing self-regulation using automated, unobtrusive assessments. They argue that it is far more important to interpret online behaviors rather than to merely count them. One way to do so is to employ a **model-tracing algorithm** that compares learners’ online help seeking behaviors to an idealized model while interacting with an intelligent tutoring system (ITS). ITSs are able to provide hints, problem-solving strategies, and feedback to learners in real time. The log of these interactions may be used to interpret a learner’s actions and intentions in a context-sensitive manner.

Aleven et al. (2010/this issue) describe an ITS system with five main components. The first is an exhaustive set of production rules that model a well-defined problem-solving sequence such as geometry or algebra problems that possess a unique solution. The second and third components are included within a **help tutor** that identifies and provides the learner either problem solving hints or feedback. The fourth component is an unobtrusive log of tutor–learner interactions. The fifth component is a set of scoring rules that can be used to interpret learners’ behaviors based on their requests for help, use of hints from the tutor, and subsequent problem-solving success. Needless to say, the system Aleven et al. describe is complex and evolving. Nevertheless, complex, context-sensitive measurement systems such as this may provide extremely useful information to learners as well as time-stamped traces of self-regulatory interactions, even if the ITS system requires further modifications. Indeed, two experiments suggest that their ITS leads to highly significant learning gains.

Aleven et al. (2010/this issue) argue convincingly that the model-tracing approach provides a helpful complement to think aloud measures because it is automated, unobtrusive, context-sensitive, and well suited for collecting data over
extended periods. Nevertheless, one question is the extent to which the model-tracing approach is truly unobtrusive, even though the information gathered using this method is no doubt invaluable to researchers. For example, in some cases, awareness of ongoing measurements may change study behaviors. A second question is whether the model-tracing approach provides the same breadth and depth of data as think alouds, feeling-of-knowing judgments, and the palette choices described in Azevedo et al. (2010/this issue). As described by Aleven et al., model-tracing is an interpretative process that matches a student’s behavior against a model embedded in the software. It is unlikely that model tracing provides specific information about strategies and the rationale for choices and decisions that think alouds provide.

Graesser and McNamara (2010/this issue) address the challenges of building an intelligent tutor that can scaffold SRL in real time. To do so, it is necessary to construct an adaptive system that can detect and respond to a learner’s ongoing needs for assistance. Three types of collaborations are discussed, including student-centered, tutor-centered, and fully interactive discussions. Potentially, any of these discussions include dialectical mechanisms that enable the learner and tutor to interact through inquiry and collaborative exchanges, question asking and answering, argumentation to resolve conflicting ideas, and Socratic dialogues. These interactions occur ideally when the learner encounters challenges in her zone of proximal development.

Graesser and McNamara (2010/this issue) devote much of their article to a discussion of tutoring, including characteristics of an effective tutor, and how to mimic these characteristics in nonhuman tutors. Like humans, cyber tutors must possess expert knowledge, automated procedural skills, conditional awareness of the current problem situation, and courteous modeling extensions include dialectical mechanisms that enable the learner and tutor to interact through inquiry and collaborative exchanges, question asking and answering, argumentation to resolve conflicting ideas, and Socratic dialogues. These interactions occur ideally when the learner encounters challenges in her zone of proximal development.

Graesser and McNamara (2010/this issue) also describe four tutoring environments that promote SRL through instruction and training. MetaTutor (Azevedo, 2005) trains learners on how to plan, monitor, and reflect on learning using 13 core strategies. iSTART (Interactive Strategy Trainer for Active Reading and Thinking) helps students become self-regulated learners by constructing self-explanations of the text using five broad strategies (McNamara, O’Reilly, Rowe, Boonthum, & Levinstein, 2007). SEEK (Source, Evidence, Explanation, and Knowledge) Web Tutor promotes critical thinking by teaching learners how to evaluate the quality and relevance of information and claims (Halpern, 2002). iDRIVE (Instruction with Deep-level Reasoning questions In Vicarious Environments) uses animated agents to model question–answer dialogues for learners (Gholson & Craig, 2006). Despite differences among the tutoring systems, each has been shown to improve skills and learning through direct instruction of skills, modeled training, and interactive dialogue.

To be effective, however, a tutor must identify situations in which a learner struggles. AutoTutor monitors learner progress using dialogue moves, which are based in part on the amount of questions posed by the learner, the number of hints and prompts by the tutor, and responses to the tutor’s queries. Information about the frequency and type of learner responses is evaluated by latent semantic analysis statistical measures to infer the self-regulatory level of the student. Although the development of sophisticated measurement systems is extremely difficult, current tutoring programs have at least two important strengths. One is that data may be collected and acted upon by the tutor in an unobtrusive manner that does not require self-monitoring or even explicit help-seeking by the student. A second strength is that the assessment, diagnosis, and response by the tutor occur in real time in a manner that promotes dialectical interaction between the learner and tutor.

Greene, Muis, and Pieschl (2010/this issue) argue that epistemic beliefs are related to all phases of SRL and are activated by local learning conditions in each phase that affect how they impact cognitive and metacognitive processes. They draw on Winne and Hadwin’s (1998) model of SRL, which postulates four phases consisting of task definition, goal setting and planning, studying tactics, and adaptation. In addition, they suggest that epistemic beliefs affect self-regulatory processes in a domain-general, domain-specific, or a task-specific manner. Greene et al. also argue that researchers can infer epistemic beliefs from the choices that learners make while engaged in a variety of CBLEs. For example, the decision to select complex tutor–learner interactions rather than drill-and-practice software options may indicate the extent to which a learner prefers complexity and incremental learning to more simple forms of learning and instruction. Although intriguing, it is not clear how researchers would conduct these analyses. One possibility is to validate the empirical choices learners make with self-reported epistemic beliefs. An alternative would be to develop a sophisticated scoring rubric for assessing the relationship between choices and beliefs. Given the development of a measurement and scoring system to evaluate the effect of tacit beliefs, such a system would provide an invaluable tool for understanding the relationships between beliefs and online processing decisions.

Greene et al. (2010/this issue) also suggest a more practically feasible strategy in which epistemic belief statements are embedded in learning materials such as an extended text to capture individuals’ beliefs about knowledge and knowing. This technique used with think alouds or pretest self-reported epistemic beliefs could be used reliably to examine the relationships between beliefs, study time, receptivity to information, drawing conclusions, and generating sophisticated arguments based on text-based evidence.
One particular methodological strength of this article is a discussion of how the four macrolevel processes described in the Winne and Hadwin (1998) model may be decomposed into a variety of microlevel indicators of self-regulation that are linked to other research of CBLEs (Azevedo, 2005; Green & Azevedo, 2009; Pintrich, 2000). Indeed, Greene and Azevedo (2009) discussed the linkage of 30 microlevel indicators to the four macrolevel stages in Winne and Hadwin (1998). From a measurement perspective, this set of 30 self-regulation measures provides a more comprehensive and accurate set of indicators when explaining moment-by-moment local as well as global, self-regulation processes.

A TAXONOMY OF MEASUREMENT STRATEGIES

The goal of this special issue is to better understand the measurement of SRL in CBLEs, with special emphasis on measuring moment-by-moment self-regulatory events compared to global measures of self-regulatory aptitudes or motivational attitudes (Zimmerman, 2008). Although a number of previous articles have discussed measurement issues in SRL (Azevedo & Witherspoon, 2009; Boekaerts & Corno, 2005; Ericsson, 2006; Nicol, 2009; Winne & Nesbit, 2009; Winne & Perry, 2000; Wolters, Pintrich, & Karabenick, 2005; Zimmerman, 2008), none of them have provided a schematic summary of measurement methods used in CBLE research. Figure 1 summarizes a variety of measurement strategies which are discussed in the four invited articles in this issue and extends Azevedo (2008, Figure 6.1). I have partitioned these measurement strategies into offline and online measures. Offline measures refer to measurements taken before or after the primary learning episode and are partitioned further into self-reported beliefs, ability, and expected performance measures. Online measures refer to measurements taken during the primary learning episode and are partitioned further into unobtrusive and obtrusive performance measures. I used the term unobtrusive to refer to indicators that can be assessed without the learner’s knowledge (e.g., reading time, eye tracking). In contrast, I use the term obtrusive measures to refer to indicators that require the learner’s conscious attention (e.g., self-reports, think alouds). These measures do not necessarily impede performance, and in fact may enhance performance (e.g., reflective think alouds). They are obtrusive to the extent that the learner is aware of them, which may potentially affect task engagement and performance (Kirk & Ashcraft, 2001).

Offline measures can be partitioned into self-reported beliefs, current abilities, and expected performance. Many different types of personal beliefs have been studied in education, including epistemological beliefs (Greene et al., 2010/this issue). Several self-report questionnaires that measure epistemological beliefs have appeared in the literature (Schommer, 1990, Schraw, Bendixen, & Dunkle, 2002). These instruments typically include 30 to 60 statements that can be completed in less than 10 min using a multipoint Likert scale. There is debate currently whether these beliefs should be assessed as domain-general versus domain-specific phenomena (Alexander, 2003). Greene et al. suggest that both approaches should be used if possible. In addition, many self-efficacy instruments have been used over the past 30 years. These measures typically are task specific in nature (e.g., assessing self-efficacy for algebra). In contrast, a variety of instruments have been developed to assess metacognition (Schraw & Dennison, 1994) and self-regulation (Pintrich, 2000) that are domain-general in nature. These measures focus on a repertoire of self-regulatory skills and motivational proclivities that affect self-regulation in general but may have no effect on specific events that occur during online self-regulation of learning. Of importance, a great deal of current research, and the articles included in this special issue in particular, focus on the measurement of specific online indices, which may or may not be correlated with self-report measures of self-regulation.

Current abilities include a wide variety of domain-general or domain-specific skills that learners bring to a learning task. Examples of domain-general skills include working memory, intelligence, reading fluency, and general reasoning skills. Examples of domain-specific skills include prior knowledge in a specific domain (e.g., biology) and pretest and posttest measures on specific problem-solving skills in any domain of interest. Many researchers have used domain-general abilities as covariates in experimental designs. In addition, domain-general abilities can be used as blocking variables to examine in greater detail the relationships, and especially attribute-treatment interactions, between different levels of the blocking ability (e.g., levels of prior knowledge and expertise) and an experimental manipulation (e.g., inclusion of hyperlinks).
Expected performance variables include judgments of learning (JOLs) made before a learning task and preactivity verbal reports of goals and action plans for performance. JOLs typically require an individual to assess some facet of learning before it occurs, even though delayed JOLs (i.e., judgment made after the learning event) are more accurate (Schraw, 2009). In contrast, verbal reports enable a learner to discuss in depth factors that affect future learning and performance such as the role that prior knowledge, ability, or beliefs might play in learning. Verbal reports also allow an individual to articulate his or her plans, goals, intended strategies, and criteria for successful learning. These measures could be used as covariates or blocking variables, or correlated with online data in the same manner as self-reported beliefs.

Online measurements occur during learning. Some of these measures are unobtrusive because the learners do not know they are being measured (e.g., reading times using computer software) or because they do not require attention or monitoring on the part of the learner (e.g., choices made regarding materials presentation format or desired software). The most general class of unobtrusive measures consist of trace logs (Winne & Hadwin, 2008; Winne & Nesbit, 2009), which include any type of data collected as a trace of some ongoing activity. Examples include reading or processing time measures, eye tracking, looking forward or backward in a document, or requesting help from an automated tutor. CBLEs enable researchers to collect many types of trace logs that can be analyzed at the microlevel (e.g., using time data to make inferences about ease of processing) or at a more macrolevel (e.g., using choices to infer strategy use or epistemological beliefs) as described by Greene et al. (2010/this issue). The main challenge associated with these types of data is to draw a valid inference that relates online, yet implicit cognitive processes, to individual traces.

Collectively, the articles also discuss three other types of unobtrusive measurements that can be used to infer self-regulatory processes. One is hyperlink maneuvers that can be used to infer need for help or a desire for additional information, which may allow researchers to make inferences about otherwise implicit processes (Greene et al., 2010/this issue) refer to as "palette choices," in which a learner is given several options about how to proceed or to select additional materials. These choices may allow researchers to make inferences about strategy use, depth of understanding, beliefs, or motivation (Greene et al., 2010/this issue). The third is inserted belief statements that are included in documents, which may elicit trace logs such as reading time or records of learner choices that can be used to make inferences about otherwise implicit processes (Greene et al., 2010/this issue). These include epistemic beliefs, coordination of multiple sources of information, and the construction and evaluation of arguments.

The four invited articles also discuss five generic obtrusive measures that under certain circumstances may require use of limited processing resources of learners, and therefore may interfere with online processing. The first is think alouds, which provide extremely rich information about online self-regulatory processes even though they may be time intensive to score. Generally, concurrent reports are viewed as more accurate and valid indicators of mental activity then retrospective reports, which may be subject to knowledge-based intrusions. A second measure is note taking using e-logs, which provides a wealth of data regarding what learners perceive to be important. The structure of notes also may reveal how deeply a learner understands information. A third strategy is to use inserted quizzes that can be administered and scored electronically and, if desired, combined with feedback to the learner. A fourth measure includes a log of learner–tutor interactions that teach strategies, provide feedback, and test as well (Aleven et al., 2010/this issue; Graesser & McNamara, 2010/this issue). This type of data is especially helpful when considering the effectiveness of competing automated tutoring programs. A fifth measure is the general class of calibration judgments that learners make online to monitor and self-regulate their performance. These include feeling-of-knowing, confidence, and performance accuracy judgments (Schraw, 2009; Zimmerman, 2008). As Graesser and McNamara argue, self-regulation may be impossible without the ability to monitor one’s learning accurately.

FOUR MEASUREMENT CHALLENGES

The articles in this volume suggest four important challenges in the measurement of self-regulation in CBLEs. The first is the challenge of automating measurements in computer environments, which include the development and implementation of pedagogical agents, scripting ongoing interactions, providing feedback to learners, and orchestrating a repertoire of automated assessments to capture these ongoing activities. The current contributions indicate that great progress has been made in terms of designing and building software to teach and tutor learners, although less progress has been made toward the goal of assessing the effects of computer learning environments. The current issue is devoted specifically to highlighting the importance of this issue for future research.

A second challenge is to create rubrics to score the complex measures shown in Figure 1. For example, although it is relatively easy to collect verbal report data before or during an intervention, it is far more difficult to create a reliable scoring rubric, or to develop online coding schemes, to accurately capture the breadth and depth of think aloud responses. To complicate matters, it is even more difficult to standardize these rubrics and coding systems across experiments and research teams to evaluate the generalizability of findings (Ericsson, 2006; Nisbett & Hadwin, 2006). Notwithstanding these challenges, researchers have created reliable scoring systems to evaluate think aloud data (Azevedo & Witherspoon, 2009). For example, Kuhn (1991) analyzed the
quality of verbal arguments in response to open-ended dilemmas in which she examined the type and number of claims, supporting evidence, and warrants for arguments. A different approach has been to use computer software such as ATLAS.ti (Muhr, 2004) to construct IF–THEN argument graphs that illustrate claims made during think alouds, evidence in support of those claims, and the direct and indirect relationships among claims (Olauson, Schraw, & VanderVeldt, 2010).

A third challenge is to interpret the constructs embedded in the measures in Figure 1 after data are collected and scored using reliable rubrics. Unfortunately, some of the constructs of interest to researchers such as epistemological beliefs (Greene et al., 2010/this issue), palette choices (Azevedo et al., 2010/this issue), mental models (Graesser & McNamara, 2010/this issue), and the quality of help seeking interactions (Alev et al., 2010/this issue) have proven very difficult to measure reliably using any of several measurement strategies (Schraw & Olauson, 2002, 2008). This is not intended as a criticism of these constructs or the current authors’ efforts but rather the difficulty in defining and accurately measuring complex, oftentimes highly elusive phenomena. The development of a taxonomy that defines constructs and links measurements to these constructs would advance CLBEs research tremendously.

A final challenge is aligning multiple sources of empirical data to validate a single construct, or more challenging still, multiple measures of different constructs to triangulate findings about a learner’s SRL in general. Both of these scenarios pose serious measurement challenges, especially when one considers that the goal of most SRL studies is to understand the cognitive activities that support self-regulation, rather than to provide evidence that self-regulation was measured reliably.

There are no easy solutions to these measurement challenges. Clearly, it behooves researchers to expand the available repertoire of measurement strategies in order to utilize multiple, convergent measures of self-regulation as well as to ensure that instruments and measurement strategies yield reliable and valid inferences. To do so, researchers must articulate the construct(s) that each measurement tool assesses, provide adequate coding and scoring schemes, and situate each measurement strategy within a broader measurement framework that expands on the taxonomy shown in Figure 1.

**SOME IMPLICATIONS FOR MEASURING SELF-REGULATED LEARNING**

The current issue makes a number of explicit and implicit suggestions for improving the measurement of SRL in CBLEs. One is to develop an integrated repertoire of measures of SRL. A variety of articles have appeared over the last decade that describe specific measures (Perry & Winne, 2006), whereas others have focused on broader measurement issues of SRL (Winne & Nesbit, 2009). The development of a wide variety of measurement strategies, both obtrusive and unobtrusive, has been one of the important contributions of CBLEs research. Figure 1 represents my preliminary attempt to integrate some of these methods into a taxonomy that ideally will promote further research and development.

Figure 1 is useful for understanding different types of SRL measures, but it does not address the relative strengths and weaknesses of each measure. Table 1 attempts to provide an initial comparison of measurement strategies that can be pursued in greater detail in the future, focusing on salient cognitive processes that are discussed within most theories of SRL. These processes include prestudy factors that could be used as covariates or blocking variables; measures of expected or online effort; and the extent to which learners generate and use plans and strategies, monitor their learning, integrate information, and show evidence of increased learning.

Table 1 indicates with an X whether each measure can be used to make a thorough assessment of the seven different cognitive processes shown in row 1. No doubt, a compelling

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>The Utility of Outcome Measures for Assessing Cognitive Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Processes</strong></td>
<td><strong>Prestudy Factors</strong></td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
</tr>
<tr>
<td>Self-reports</td>
<td>X</td>
</tr>
<tr>
<td>Cognitive abilities</td>
<td>X</td>
</tr>
<tr>
<td>Expected performance</td>
<td>X</td>
</tr>
<tr>
<td>Processing time</td>
<td></td>
</tr>
<tr>
<td>Eye tracking</td>
<td>X</td>
</tr>
<tr>
<td>Hyperlink choices</td>
<td></td>
</tr>
<tr>
<td>Think alouds</td>
<td>X</td>
</tr>
<tr>
<td>Inserted quizzes</td>
<td></td>
</tr>
<tr>
<td>Palette choices</td>
<td></td>
</tr>
<tr>
<td>Auto-tutoring</td>
<td></td>
</tr>
<tr>
<td>Calibration judgments</td>
<td></td>
</tr>
</tbody>
</table>
argument could be made by CBLEs experts for or against each cell in Table 1; however, the main goal of my informal summary is to suggest that a more detailed summary matrix that expands on Table 1 would be useful for the field in general, and the study of microlevel CBLEs in particular. To advance this research, the CBLEs community needs to identify salient self-regulatory skills and processes—and cross them with a variety of measurement strategies—that enable researchers to fully understand the processes and final performance-based products that constitute self-regulation. Experts should consider carefully which self-regulatory processes are most important to their research and determine which type of measurement strategy is a valid and useful measure of that process. Distinguishing between skills at different levels of complexity is important too.

Two measurement-related conclusions seem justified based on Table 1. One is that each of the seven cognitive processes can be assessed by at least two or more types of measures. For example, online monitoring can be assessed using traditional measures such as feeling-of-knowing judgments, as well as think alouds and learner–tutor traces. Other measures in Table 1, or measures not shown in Table 1, might be used to assess online monitoring as well under certain experimental conditions. Fortunately, this suggests that it is feasible (and beneficial) to collect multiple, convergent measures for crucial self-regulatory processes.

A second conclusion is that each measure can be used to assess two or more cognitive processes. Thus, researchers may use the same measure to make inferences about multiple cognitive processes. The utility of each measure probably will depend on the cognitive processes being studied and the experimental manipulations of interest; however, it behooves researchers to plan in advance ways to maximize the amount of information collected in order to maximize the conclusions that can be drawn about targeted cognitive processes.

Table 1 also suggests that multiple measures can be combined to gather convergent evidence of each type of cognitive process. I encourage researchers to do so to develop what has been known historically as a nometheic net (Shadish, Cook, & Campbell, 2002) that integrates theory and measurement into a single conceptual framework that can be evaluated using a multiple-trait, multiple-method validation strategy that uses multiple traits (e.g., strategy use, planning, monitoring, interest, self-efficacy) and multiple methods (e.g., self-report and think alouds). For example, online strategy use could be evaluated based on think aloud reports, eye-tracking data that supports or fails to support intended strategy use described before the study, interactions with tutors, as well as palette choices made in real time. Using multiple measures and methods should enhance the interpretative validity of conclusions by combining and synthesizing multiple sources of data and allowing researchers to examine the concurrent validity (i.e., real-time relationships) among different measures.

Multiple measures could be used profitably in data analyses as well. One option is to compare treatment and nontreatment groups using multiple measures entered into a discriminant function analysis to determine which of the measures best discriminates between groups. A second option is to use these measures in some type of multiple regression analysis to determine the amount of variation each measure explains with respect to a specific learning outcome.

Multiple measures could be especially helpful when used to create a structural equation model of the hypothesized interrelationships among crucial cognitive processes such as integration of meaning and construction of mental models (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Graesser, Jeon, & Dufty, 2008). These models could specify the causal relationships between salient steps in the self-regulation process as suggested in recent theories (Azvedo et al., 2010/this issue; Graesser & McNamara, 2010/this issue; Winne & Hadwin, 2008).

TWO UNRESOLVED VALIDITY ISSUES

Validity is the extent to which evidence supports inferences and conclusions about the phenomenon of interest. It is a truism in educational research that more effort and rigor should be applied to documenting the validity of tests and measures used to understand and evaluate performance. However, two issues seem particularly important regarding the measurement of CBLEs and the effectiveness of experimental interventions that compare CBLEs. One is the extent to which obtrusive measures interfere with processing by biasing responses in a certain direction. For example, asking learners to rate their epistemological beliefs before the study, or to think aloud about these beliefs during the study, may change the way they engage in the task for better or for worse. In addition, collecting reading time data may alter cognitive processing. More attention should be paid to the possibility that self-awareness of specific measures may create response bias when evaluating CBLEs (Hertzum et al., 2009).

A second issue is the extent to which measures in Figure 1 compete for the learner’s limited cognitive resources. Previous research has investigated the effect of think alouds on online performance and concluded that verbal reports can be used in many settings without disrupting processing, and without compromising the validity of data (Ericsson, 2006). This especially is true of concurrent compared to retrospective verbal reports. However, many of these studies were conducted with experts; thus, it is less clear how nonexperts perform, or how complex CBLEs affect either experts or nonexperts. It may be the case that think alouds are more useful for some types of activities than others, in part due to explicit awareness of these processes as well as the complexity of the process that competes with self-reporting.
SUMMARY

Understanding how computer and classroom technologies influence and improve learning is of utmost importance to educators. Previous research suggests that CBLEs improve higher level thinking skills and deep comprehension (Azevedo & Witherspoon, 2009; Winne & Hadwin, 2008; Winne & Nesbit, 2009). Nevertheless, development and validation of an integrated measurement system that assesses multiple traits using multiple methods across multiple settings lags behind the development of the CBLEs technologies. Researchers are faced with the ongoing challenge of measuring the effects on learning in a reliable and valid manner. I have argued here that to do so, researchers must develop a systematic, integrated repertoire of measures of SRL in CBLEs. Ideally, these measures should be shown to be reliable and valid, be aligned to one another concurrently, be exhaustive in the sense that they measure all cognitive processes of interest, and enable the researcher to use multiple outcome measures (e.g., beliefs, processing time measures, eye tracking, and think alouds) to triangulate each type of learning. Significant progress has been made in this direction, as the present issue indicates. However, future work should focus on developing an integrated framework for understanding the deployment and measurement of CBLEs. Doing so will allow researchers to establish convergent evidence in support of CBLE technologies.

REFERENCES


