Assessing 21st Century Skills: Integrating Research Findings

National Council on Measurement in Education
Vancouver, B.C.

Emily R. Lai
Michaela Viering

April 2012

Abstract

This paper synthesizes research evidence pertaining to several so-called 21st century skills: critical thinking, creativity, collaboration, metacognition, and motivation. We provide a rationale for focusing on these five skills and explain their importance as educational outcomes. We then summarize peer-reviewed research published in education and psychology in order to answer several questions about these skills: (1) how do researchers define them; (2) how are they related to one another both theoretically and empirically; and (3) how do researchers traditionally measure them. We use the answers to these questions to make several recommendations regarding how best to assess these skills. Briefly, research suggests these skills are inter-related in complex ways. Researchers have used a number of approaches to measuring these skills, including (1) self-reports, (2) global rating scales, (3) standardized assessments, both multiple-choice and performance-based, and (4) observational measures. We recommend several practices for assessing 21st century skills: incorporating multiple measures to permit triangulation of inferences; designing complex and/or challenging tasks; including open-ended and/or ill-structured tasks; using tasks that employ meaningful or authentic, real-world problem contexts; making student thinking and reasoning visible; and exploring innovative approaches that utilize new technology and psychometric models.

Keywords: 21st century skills, college readiness, critical thinking, creativity, collaboration, metacognition, motivation, performance-based assessment

Assessing 21st century skills: Integrating research findings

Background

"21st century skills" and "college and career readiness" have recently become watchwords in education. The Partnership for 21st century skills advocates adoption of local, state, and federal policies that support explicit integration of 21st century skills into instruction for all students (P21, 2009). The two consortia formed in response to Race to the Top both highlight college and career readiness as their primary educational target (PARCC, 2010; SBAC, 2010). Furthermore, the Common Core State Standards anchor K-12 academic standards in expectations that all students will be college-or career-ready upon high school graduation (CCSS, 2010).

Along with new emphases on including 21st century skills in curricula and instruction is a growing need to assess students' competency at these skills on a large scale. Accordingly, recent policy initiatives have contributed to conversations about attributes of so-called "next generation" assessment systems. For example, according to the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium, assessment tasks must "measure rigorous content and students' ability to apply that content." Such tasks will "elicit complex demonstrations of learning and measure the full range of knowledge and skills necessary to succeed in college and 21st-century careers." Additionally, these tasks will "send a strong, clear signal to educators about the kinds of instruction and types of performances needed for students to demonstrate college and career readiness" (PARCC, 2010, p. 35).

Which 21st Century Skills?

Although these skills are not new, it was not until very recently that educators and policy makers agreed that they should be explicitly included in academic content standards, directly taught alongside the regular academic curriculum, and routinely assessed for all students. Despite widespread agreement on their importance, however, there still appears to be disagreement as to which particular skills matter for college and career readiness. Numerous skills frameworks exist and different frameworks identify different skills as important. For example, the Common Core State Standards (CCSS) in English Language Arts specifically call for instructional emphasis on "application of knowledge through higher-order skills," such as the ability to create and support arguments based on evidence and logical reasoning through writing and sharing ideas with classmates via speaking and listening during informal collaboration. Similarly, the CCSS in Mathematics highlight student ability to "practice applying mathematical ways of thinking to real world issues and challenges."

The Partnership for 21st Century Skills (P21) has created a comprehensive framework for conceptualizing different types of skills important for college and the workforce. For example, learning and innovation skills include creativity and innovation, critical thinking and problem solving, and communication and collaboration. Information, media, and technology skills include information literacy, media literacy, and information/communications/technology literacy. Finally, life and career skills include flexibility and adaptability, initiative and self-direction, social and cross-cultural skills, productivity and accountability, and leadership and responsibility (Partnership for 21st Century Skills, 2009).

The National Research Council initiated an investigation into the topic of teaching and assessing 21st century skills, hosting several workshops and seminars beginning in 2005. Out of this work came a framework for categorizing the types of knowledge and skills students need for college and career readiness: (1) cognitive skills, including critical thinking, non-routine problem solving, and systems thinking; (2) interpersonal skills, including complex communication, social skills, teamwork, cultural sensitivity, and dealing with diversity; and (3) intrapersonal skills, including self-management, time management, self-development, self-regulation, adaptability, and executive functioning (Committee on the Assessment of 21st Century Skills, 2011).

Finally, the Assessment and Teaching of 21st Century Skills (ATC 21) organization has also offered a framework for organizing different types of 21st century skills (Binkley et al., 2010). This framework includes four classes of skills:

1. Ways of Thinking, which encompasses creativity and innovation; critical thinking, problem solving, and decision-making; and metacognition or learning to learn
2. Ways of Working, which includes communication and collaboration or teamwork
3. Tools for Working, which addresses information literacy and information and communication technology (ICT) literacy
4. Living in the World, which includes citizenship, life and career skills, and personal and social responsibility

The ATC 21 framework identifies several important dimensions that cut across these different types of skills, referred to as the KSAVE model, where the acronym represents Knowledge, Skills, and Attitudes/Values/Ethics. This model suggests that there are knowledge components, skill components, and affective, dispositional, or attitudinal components within each of the four core skills types.

There are clearly areas of overlap among these various frameworks. For example, the NRC cognitive skills category appears to be roughly equivalent to P21's learning and innovation category and to ATC 21's "ways of thinking" category. Similarly, NRC's intrapersonal skills category appears to be the analog to P21's life and career skills category and also somewhat comparable to ATC 21's "living in the world" category. Finally, P21's information, media, and technology skills category is directly comparable to ATC 21's "tools for working."

There is also overlap at the level of individual sub-skills. Table 1 reports these areas of overlap among the P21, NRC, and ATC 21 frameworks, as well as links to recognizable constructs with relatively mature research bases. For example, critical thinking and collaboration emerge as important skills in all three frameworks, although the P21 considers collaboration to be a learning and innovation skill, the NRC classifies it as an interpersonal skill, and ATC 21 identifies collaboration as a "way of working." Creativity is specifically identified within the P21 and the ATC 21 frameworks, and is also evident in the NRC's emphasis on non-routine problem solving. Motivation is highlighted in P21's focus on the life and career skill of "initiative" as (1) exploring and expanding one's learning and opportunities to gain expertise; (2) holding a commitment to lifelong learning; and (3) completing tasks independently. Motivation is also implied in P21's definition of "flexibility," particularly "dealing positively with praise, setbacks, and criticism." The NRC framework does not explicitly call out motivation, but it is invoked within the intrapersonal skills category via "self-development" and "adaptability." Similarly, the ATC 21 framework reflects an implied emphasis on motivation within the "living in the world," category, particularly the life and career skills of adaptability, flexibility, and self-direction as "going beyond basic mastery," and "demonstrating initiative." Finally, metacognition is specifically called out in the ATC 21 framework as a "way of thinking." It is also identified as a life and career skill within the P21 framework in the form of self-direction and productivity, whereas it is evident within the NRC framework as the intrapersonal skills of self-management and self-regulation.

Although there are other apparent areas of overlap in some of these frameworks—for instance, ICT literacy—the constructs represented in Table 1 are those that correspond to a relatively robust and established research basis within education and psychology. Thus, these are the constructs we chose to focus on.

Table 1. Mapping of 21st Century Skills Frameworks

Research-based construct	P21 Framework terminology	NRC Framework terminology	ATC 21 Framework terminology
Critical thinking	Learning and innovation – critical thinking	Cognitive – critical thinking	Ways of thinking – critical thinking, problem-solving, and decision-making
Collaboration	Learning and innovation – communication and collaboration	Interpersonal – complex communication, social skills, teamwork	Ways of working – communication and collaboration
Creativity	Learning and innovation – creativity and innovation	Cognitive – non-routine problem solving	Ways of thinking – creativity and innovation
Motivation	Life and career skills –initiative, flexibility	Intrapersonal – self-development, adaptability	Living in the world –adaptability, flexibility, self-direction
Metacognition	Life and career skills – self-direction, productivity	Intrapersonal - self-management, self-regulation	Ways of thinking – metacognition or learning to learn

Why Are They Important?

Educators have long touted the importance of developing critical thinking skills as an end in itself. Moreover, critical thinking skills have been demonstrated to predict important educational and employment outcomes in a number of settings (see, for example, Miller, Sadler, & Mohl, 1993). Similarly, motivation researchers argue that encouraging motivation in children is critical because it predicts motivation later in life (Broussard & Garrison, 2004; Gottfried, 1990). Motivation is also related to achievement and IQ. Research demonstrates a relatively consistent relationship between motivation and achievement in reading and math (Broussard & Garrison, 2004; Gottfried, 1990; Lange & Adler, 1997). Intrinsically motivated first-grade students tend to have higher achievement in these subjects than extrinsically motivated students, and mastery (or intrinsic) motivation predicts later reading and math achievement, whereas judgment (or extrinsic) motivation does not. Indeed, Lange and Adler (1997) found that motivation contributes to the prediction of achievement over and above ability.

Interventions targeted at improving creative thinking have also been successful at increasing student academic achievement (Maker, 2004). Similarly, studies have shown that measures of creative thinking significantly predict first-year college students' grade point averages (GPAs) above and beyond high school GPA and SAT scores (Sternberg, 2006b), as well as success in graduate school (Frederickson & Ward, 1978). Lubart and Guignard (2004) argue more generally that as technology continues to advance, people will increasingly be required to think in creative and divergent ways in order to address new types of problems. In turn, creativity itself precipitates additional societal and technological changes in that it drives the development of new ideas, inventions, and technologies. New approaches to fields of study lead to innovations that move the fields forward—either by looking at old ideas in new ways, advancing current thoughts, introducing completely new concepts, or by integrating diverse concepts in new ways (Lubart & Guignard, 2004).

Metacognition may help to compensate for deficits in intelligence or prior knowledge of a subject during problem solving (Prins, Veenman, & Elshout, 2006). For example, as demonstrated in research studies, students with high metacognitive skill levels tend to outperform students with low metacognitive skills on complex and unfamiliar tasks, even when the two groups are equally matched in ability or aptitude. Some researchers speculate that this is because students with strong metacognitive skills activate problem-solving heuristics (such as creating a graphical representation of a word problem) and "improvise" strategies more efficiently than students without such metacognitive skills (Prins et al., 2006).

Collaboration also has powerful effects on student learning. These effects are seen in the form of higher scores on work completed collaboratively, even when students turn in separate products (Fall, Webb, & Chudowsky, 1997; Rojas-Drummond & Mercer, 2003; Saner, et al., 1994; Webb, 1993). In addition, evidence from these studies suggests that learning that occurs during collaboration persists (Saner et al., 1994; Webb, 1993). In other words, after collaborating with others, a student's performance on subsequent, related tasks completed individually tends to be higher than the performance of similar-ability students who only work alone. Thus, engaging in collaborative learning opportunities with classmates can have a lasting impact on individual student learning. Collaborating can also increase students' social competency (e.g., conflict resolution skills and use of helping behaviors) and academic self-concept (Ginsburg-Block, Rohrbeck, & Fantuzzo, 2006).

Research Objectives

We sought to tie current conversations around teaching and assessing 21st century skills to pertinent research literature within educational psychology. In particular, we sought to answer the following questions:

1. How do researchers define these constructs?
2. How are these constructs theoretically and empirically related?
3. How do researchers traditionally measure these constructs?

Our goal was to synthesize research findings in order to provide a coherent set of recommendations for building assessments that are ideally suited to measuring constructs underlying 21st century skills.

Methods

We reviewed the academic research literature pertaining to several cognitive and noncognitive constructs commonly reflected in frameworks on 21st century skills: critical thinking, creativity, collaboration, motivation, and metacognition. We focused on research addressing these skills in school-age children, as opposed to college students or adults. To identify pertinent research literature on these constructs, we worked with a Research Librarian and used resources associated with two major research university libraries. Our literature searches began in November 2010 and continued in earnest through January 2011. However, we continued to identify and acquire additional studies sporadically throughout 2011 as our work progressed.

Research informing these constructs comes from diverse fields, including educational psychology, educational philosophy, computational linguistics, and gifted and talented education. Given the multidisciplinary nature of these 21st century skills, we initially cast a wide net. We began by using Google Scholar to conduct broad searches, using several keywords associated with these constructs. For example, for metacognition, we searched for publications linked to “metacognition,” “self-regulation,” and “self-regulated learning.” We read abstracts and titles to determine each item’s relevance to the research questions identified above.

We then conducted more targeted searches within specific journals and electronic databases, based on our initial broad search. For example, we searched by keyword within EBSCOhost, JSTOR, Web of Science, Education Full Text, and APA PsychNET. Similarly, we searched by keyword within specific journals likely to identify prior literature reviews of the topics, such as the Review of Educational Research and Educational Research Review.

Finally, we conducted “snowball sampling” techniques, whereby we culled the references sections of relevant studies to identify additional, relevant studies. We also consulted with colleagues in outside organizations and universities who are experts in educational and cognitive psychology for additional recommendations. Our goal was to achieve a balance of research studies that represented both seminal (but possibly dated) works and more recent publications. We included some previous literature reviews, but preferred to cite original research to the extent possible.

Ultimately, we based our reviews on approximately 150 studies, with between 25 and 44 studies supporting each of the separate skills we researched (i.e., critical thinking, creativity, metacognition, etc.). For each of these skills, we wrote a separate literature review focused on answering the research questions identified above, among others. We then synthesized the research findings across the five reviews to identify a set of assessment characteristics that support their measurement, such as item type, task structure, response mode, and distribution across content areas.

Results

How Do Researchers Define 21st Century Skills?

Critical Thinking. Educators have long been aware of the importance of critical thinking skills as an outcome of student learning. Despite widespread recognition of its importance, there is a notable lack of consensus regarding the definition of critical thinking. Sternberg (1986) identified several schools of thought on the issue, with one school represented by cognitive psychologists and the other represented by philosophers of education. These two approaches differ primarily in terms of two dimensions: (1) whether researchers focus on the ideal critical thinker versus how people actually think and (2) whether they focus on qualities of thought versus actions or behaviors exhibited by critical thinkers (with the former characterizing the philosophical approach and the latter representing the cognitive psychological approach). One classic definition of critical thinking developed in the philosophical tradition depicts it as "reflective and reasonable thinking that is focused on deciding what to believe or do" (Ennis, 1985, p. 45). A corresponding popular definition offered by cognitive psychologists is that critical thinking is "the use of those cognitive skills or strategies that increase the probability of a desirable outcome" (Halpern, 1998, p. 450). More recently, critical thinking has been compared to rational thinking, defined as "adopting appropriate goals, taking the appropriate action given one's goals and beliefs, and holding beliefs about the world that are commensurate with available evidence" (West, Toplak, & Stanovich, 2008, p. 931).

Despite differences among the schools of thought and their approaches to defining critical thinking, there exist areas for agreement. First, researchers of critical thinking typically agree on the specific subskills encompassed by the definition, which include:

- analyzing arguments, claims, or evidence (Ennis, 1985; Facione, 1990; Halpern, 1998; Paul, 1992);
- making inferences using inductive or deductive reasoning (Ennis, 1985; Facione, 1990; Paul, 1992; Willingham, 2007);
- judging or evaluating (Ennis, 1985; Facione, 1990; Lipman, 1988; Tindal & Nolet, 1995); and
- making decisions or solving problems (Ennis, 1985; Halpern, 1998; Willingham, 2007).

Other abilities or behaviors identified as relevant to critical thinking include asking and answering questions for clarification (Ennis, 1985); defining terms (Ennis, 1985); identifying assumptions (Ennis, 1985; Paul, 1992); interpreting and explaining (Facione, 1990); reasoning verbally, especially in relation to concepts of likelihood and uncertainty (Halpern, 1998); predicting (Tindal & Nolet, 1995); and seeing both sides of an issue (Willingham, 2007).

Most researchers also agree that in addition to skills or abilities, critical thinking entails dispositions (Facione, 1990; Toplak & Stanovich, 2002). As early as 1985, researchers working in the area of critical thinking recognized that the ability to think critically is distinct from the disposition to do so (Ennis, 1985). Empirical evidence appears to confirm the notion that critical thinking abilities and dispositions are, in fact, separate entities (Facione, 2000; Toplak & Stanovich, 2002). These dispositions have variously been cast as attitudes or habits of mind. Facione (2000) defines critical thinking dispositions as "consistent internal motivations to act toward or respond to persons, events, or circumstances in habitual, yet potentially malleable ways" (p. 64). Researchers tend to identify similar sets of dispositions as relevant to critical thinking. For example, the most commonly cited critical thinking dispositions include:

- open- or fair-mindedness (Bailin et al., 1999; Ennis, 1985; Facione 1990, 2000; Halpern, 1998);
- the propensity to seek reason (Bailin et al., 1999; Ennis, 1985; Paul, 1992);
- inquisitiveness (Bailin et al., 1999; Facione, 1990, 2000);
- the desire to be well-informed (Ennis, 1985; Facione, 1990);
- flexibility (Facione, 1990; Halpern, 1998); and
- respect for, and willingness to entertain, others' viewpoints (Bailin et al., 1999; Facione, 1990).

Finally, most researchers working in the area of critical thinking agree on the important role of background knowledge. In particular, most researchers see background knowledge as essential if students are to demonstrate their critical thinking skills (Case, 2005; Kennedy et al., 1991; Willingham, 2007). As McPeck (1990) has noted, to think critically, students need something to think critically about. Similarly, Bailin et al. (1999) argue that domain-specific knowledge is indispensable to critical thinking because the kinds of explanations, evaluations, and evidence that are most highly valued vary from one domain to another.

Some researchers go even further to argue that critical thinking is entirely domain-specific. In other words, there is no such thing as generalized critical thinking skills. For example, Kuncel (2011) argues that generalized measures of critical thinking are indistinguishable from general intelligence or ability, and that such measures do not contribute usefully above general intelligence to the prediction of important educational outcomes. However, Toplak and Stanovich (2002) demonstrate that critical thinking dispositions, such as need for cognition and reflectivity, are better predictors of performance on domain-neutral disjunctive reasoning tasks than general cognitive ability. Similarly, West, Toplak and Stanovich (2008) demonstrate that a generalized measure of critical thinking—performance on syllogistic reasoning tasks—is a better predictor of the ability to avoid cognitive biases than is general cognitive ability. Other researchers argue that critical thinking encompasses both general and domain-specific aspects (Ennis, 1985; Paul, 1992; Smith, 2002). In other words, some critical thinking skills apply to multiple domains (e.g., formal rules of logic), whereas others are unique to specific subject areas (e.g., the use of proofs in mathematics or the scientific method in science).

Creativity. Psychologists such as Vygotsky and Guilford have long maintained the importance of fostering creative development in children in order to prepare them for a changing future (Beghetto, 2010; Guilford, 1950). However, many educators still consider nurturing creativity within the purview of gifted and talented teachers, and separate from mainstream classroom curricula. This perception is beginning to change among researchers and educational policy makers who recognize the link between creativity and economic and cultural prosperity (Beghetto, 2010; P21, 2009).

As with critical thinking, there is no single accepted definition of creativity. However, the definition of creativity as "the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" encompasses much of the current thinking in the field (Plucker, Beghetto, & Dow, 2004, p. 90, emphasis in original). Underlying almost all definitions of creativity is the notion that a creative product is both novel and has some sort of value. A novel product that does not take into account task parameters (i.e., is not useful) may be considered bizarre rather than truly creative (Lubart & Georgsdottir, 2004; Lubart and Guignard, 2004). Most researchers also recognize that creative achievement requires a combination of cognitive skills, dispositions, and environmental factors (Sternberg, 2006a; Torrance, 1977; Treffinger, et al., 2002; Van Tassel-Baska & MacFarlane, 2009). Commonly recognized creativity-related cognitive skills include the ability to:

- identify problems (Sternberg, 2010; Torrance, 1977);
- generate ideas, often by thinking divergently using fluency, flexibility, originality, and elaboration(Treffinger et al., 2002; Van Tassel-Baska & MacFarlane, 2009); and
- solve problems (Torrance, 1977).

Just as important to the creative process are analytic and evaluative skills. These critical thinking skills are essential in judging whether a creative output is worth pursuing (i.e., has value) (Sternberg, 2006a).

Possessing the requisite cognitive skills does not mean a person will undertake creative endeavors (Guilford, 1950, Sternberg, 2006a). In fact, Sternberg (2006a) asserts that being creative involves a deliberate choice. Sternberg equates creativity with a habit, stating "[creative people] habitually respond to problems in fresh and novel ways, rather than allowing themselves to respond mindlessly and automatically" (Sternberg, 2010. p. 394). Creative people tend to be:

- motivated, often intrinsically (Russ, 1996; Sternberg, 2010; Van Tassel-Baska & Macfarlane, 2009)
- willing to take intellectual risks, such as sharing tentative ideas, asking questions, and attempting to do and learn new things that place themselves at risk of making mistakes or appearing less competent than others (Beghetto, 2009; Russ, 1996; Sternberg, 2010; Treffinger et al., 2002);
- open to new ideas (Russ, 1996; Sternberg, 2010; Treffinger et al., 2002; Van Tassel-Baska & Macfarlane, 2009); and
- tolerant of ambiguity (Russ, 1996; Sternberg, 2010; Van Tassel-Baska & Macfarlane, 2009).

Creative people also tend to have high creative self-efficacy or belief in their ability to generate new and meaningful ideas (Beghetto, 2010; Russ, 1996; Sternberg, 2010). Of these dispositions, motivation, intellectual risk taking, and creative self-efficacy can be especially important in the face of the natural resistance society often displays toward creative ideas that are new and untested (Sternberg, 2006a).

Creativity researchers note that environmental and cultural factors play a role in both determining an individual's creative potential, as well as interpreting creative outputs. Environmental factors that support creativity include external support or acceptance of novel ideas, products, or ways of doing things (Runco, 2004; Sternberg, 2006a) or forums in which people are exposed to or can express creative products (Lubart & Guignard, 2004; Sternberg, 2006a). In addition, the culture to which a person belongs can influence the settings and formats in which it is acceptable to be creative. For example, creativity tends to be more product-oriented in Western cultures and process-oriented in Eastern cultures (Lubart and Georgsdottir, 2004).

One issue of debate within the field of creativity is that of domain specificity. Many scholars identify domain-specific knowledge and skills as key components of creativity ((Treffinger et al, 2002; VanTassel-Baska & MacFarlane, 2009). For example, in a study of 109 second graders, Han and Marvin (2002) found that domain-general divergent thinking skills did not predict creative performance in language, math, and art. It should be mentioned, however, that divergent thinking skills are not equivalent to creativity, but are rather a subset of creativity-related cognitive skills. It is likely that creativity entails both domain-specific and domain-general components (Lubart & Guignard, 2004; Treffinger et al, 2002). Creative outputs (products, performances, and ideas) often display domain specificity dependent on the knowledge and skill sets unique to the field in which a particular creative output is generated. However, some of the creativity-related skills and dispositions used to reach that output may be generalizable, such as the ability to combine seemingly disparate ideas in unique ways (Lubart & Guignard, 2004). Although a certain level of domain-specific knowledge is essential for creativity, too much knowledge may actually hinder creativity. A person may have difficulty manipulating the concepts within a particular field in novel ways if he or she is too ingrained in traditional thinking (Sternberg, 2010).

Finally, two related issues of debate among scholars are what constitutes a creative output and who can be creative. Although there is a general consensus that creative products are both novel and serve a purpose, there is disagreement as to what level of novelty is required to make something creative. Some scholars argue that in order for something to be creative, it must have an impact on society as a whole. In this eminent view of creativity, very few ideas or products are truly creative. This view implies that only a small number of highly talented individuals are actually creative (Van Tassel-Baska & MacFarlane, 2009).

Runco (2003), Treffinger (2002), and Lubart (2004) are among the scholars who recognize a much broader range of creative products: a creative idea or output may be novel only for the individual who produces it, it may be novel for a particular peer group or field, or novel for society as a whole. What varies is the overall level of impact of the idea or creative output (e.g. individual level or societal level) (Lubart & Guignard, 2004; Runco, 2003; Treffinger et al., 2002). In this view, everyone has the potential to be creative. This personal level of creativity in which something is novel only for an individual is sometimes referred to as "everyday creativity" (Runco, 2003) or "subjective creativity" (Kaufman, 2003).

Collaboration. Educators in a variety of educational settings—from K12 to the university classroom—have long used collaborative approaches to teaching and assessing students. More recently, educators and policy makers have identified the ability to collaborate as an important outcome in its own right rather than merely a means to teach and assess traditional academic content. Roschelle and Teasley (1995) define collaboration as a "coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (p. 70). They define the joint problem space as the shared knowledge structure that supports problem-solving by integrating goals, descriptions of the current problem state, and awareness of potential strategies, as well as the links between these things. According to Roschelle and Teasley, collaboration takes place within this joint problem space, which provides the structure needed to allow meaningful conversations about the problem. To construct a joint problem space, partners must have ways to introduce and accept knowledge, monitor exchanges for evidence of divergent meanings, and repair any divergences identified.

Blatchford et al. (2003) discuss student collaboration in the context of group work, which they define as “pupils working together as a group or team. The teacher may be involved at various stages but the particular feature of group work—perhaps its defining characteristic—is that the balance of ownership and control of the work shifts toward the pupils themselves. Group work should involve children as co-learners…not just one pupil helping another” (p. 155). Collaborative learning is broadly defined as “a situation in which two or more people learn or attempt to learn something together,” and more specifically as joint problem solving (Dillenbourg, 1999, p. 1). As Dekker et al. (2006) note, “in collaborative learning, therefore, the focus of the analysis is on the students’ effort to show and explain ideas and action to each other with the aim to find common ground for building shared understanding and strategies” (p. 66). Similarly, Gillies (2003) explains, “When children work cooperatively together, they learn to give and receive help, share ideas and listen to other students’ perspectives, seek new ways of clarifying differences and resolving problems, and construct new understandings and learning from engaging in these processes” (p. 137).

As Van Boxtel, et al. (2000) explain, collaborative learning activities allow students to provide explanations of their understanding, which can help students elaborate and reorganize their knowledge. Social interaction stimulates elaboration of conceptual knowledge as group mates attempt to make themselves understood, and research demonstrates that providing elaborated explanations improves student comprehension of concepts. Once conceptual understandings are made visible through verbal exchange, students can negotiate meaning to arrive at convergence, or shared understanding.

Collaboration is sometimes distinguished from cooperative learning in that cooperation is typically accomplished through the division of labor, with each person responsible for some portion of the problem solving. Collaboration, on the other hand, involves participants working together on the same task, rather than in parallel on separate portions of the task. However, Dillenbourg et al. (1996) note that some spontaneous division of labor may occur during collaboration. Thus, the distinction between the two is not necessarily clear-cut. According to Dillenbourg et al. (1996), in cooperation, the task is split hierarchically into independent sub-tasks and coordination is only required for "assembling partial results." Collaboration, on the other hand, may divide cognitive processes into intertwined layers, but coordination occurs throughout.

As Dillenbourg (1999) notes, there are several qualities that characterize truly collaborative interactions. For example, one marker of true collaboration is the quality of interactions, especially the degree of interactivity and negotiability (Dillenbourg, 1999). Interactivity refers to the extent to which interactions influence participants' thinking. Negotiability refers to the extent to which no single group member can impose his view unilaterally on all others, but rather all group members must work toward common understanding. Collaboration sub-skills include coordination, communication, conflict resolution, decision making, problem solving, and negotiation (Blatchford et al., 2003; Fall et al., 1997; Webb, 1995; Webb & Mastergeorge, 2003). Particularly desirable forms of communication during collaborative learning include providing elaborated explanations to teammates, asking direct and specific questions, and responding appropriately to requests from others (Gillies, 2003; Kouros & Abrami, 2006; Webb, 1995; Webb et al., 2001).

Motivation. The topic of motivation enjoys a substantial research base. Motivation is defined as all the reasons that underlie willing and volitional behavior (Guay et al., 2010). Researchers often distinguish intrinsic from extrinsic motivation. Intrinsic motivation is motivation that is animated by personal enjoyment, interest, or pleasure. As Deci et al. (1999) observe, “intrinsic motivation energizes and sustains activities through the spontaneous satisfactions inherent in effective volitional action. It is manifest in behaviors such as play, exploration, and challenge seeking that people often do for external rewards” (p. 658).

On the other hand, when people are extrinsically motivated, they pursue activities because of a desire to earn or avoid external rewards or sanctions (Ryan & Deci 2000). Educators have traditionally considered intrinsic motivation to be more desirable and to result in better learning outcomes than extrinsic motivation. Some research has demonstrated that intrinsic motivation leads to greater persistence at challenging tasks and better learning over time than extrinsic motivation (Ryan et al., 1990).

Motivation involves a constellation of one’s beliefs, perceptions, values, interests, and actions that are all closely related. As a result, various approaches to motivation can focus on cognitive behaviors (such as monitoring and strategy use), non-cognitive aspects (such as perceptions, beliefs, and attitudes), or both. For example, Gottfried (1990) defines academic motivation as "enjoyment of school learning characterized by a mastery orientation; curiosity; persistence; task-endogeny; and the learning of challenging, difficult, and novel tasks" (p. 525). On the other hand, Turner (1995) considers motivation to be synonymous with cognitive engagement, which he defines as "voluntary uses of high-level self-regulated learning strategies, such as paying attention, connection, planning, and monitoring" (p. 413).

Perceptions. Self-efficacy is one's perceived competence in a given area, as well as the perception that one is in control of his or her own performance (Bandura, 1993; Eccles & Wigfield, 2002). According to Bandura's (1982) self-efficacy theory, efficacy is the major determinant of effort, persistence, and goal setting. Empirical research supports this notion, suggesting that individuals with higher self-efficacy tend to be more motivated and successful on a given task (Pintrich & DeGroot, 1990). Self-efficacy has also been associated with the use of cognitive strategies, and self-efficacy perceptions predict achievement over and above actual ability levels (Pintrich & DeGroot, 1990).

Another line of inquiry in the field of motivation explores the issue of locus of control. According to this theory, individuals should be more motivated to the extent that they feel they are in control of their own successes and failures (Eccles & Wigfield, 2002). A person's perceptions of control over his or her own successes and failures are known as attributions (Weiner, 2000). Certain types of attributions are more likely to stimulate motivation than others. For example, empirical research suggests that those holding effort attributions tend to exhibit more positive learning behaviors, such as goal-setting that focuses on learning rather than performance (Miller & Meece, 1997), use of strategies, and persistence at difficult or challenging tasks (Stipek, 1996). When a person fails at a task, it is more motivating to attribute that failure to lack of effort than to lack of ability because the former attribution implies that with sufficient effort, the person is capable of performing the task in the future (Weiner, 2000). On the other hand, when a person succeeds at a task, it is more motivating to attribute that success to strong ability than to strong effort, because it enhances self-efficacy (Schunk, 1983).

Values. Values are incentives for engaging in certain activities. People hold different kinds of values, and these different kinds of values sustain student motivation more or less effectively (Eccles & Wigfield, 2002; Stipek, 1996). For example, when students intrinsically value a skill, they personally enjoy performing that skill regardless of whether or not they are successful at it. Students holding such values are more likely to persist at challenging tasks than students who simply value successful performance (i.e., students who hold attainment values) or who only reflect on the negative aspects of performing the skill, such as anxiety and fear of failure (i.e., students who hold cost values) (Eccles & Wigfield, 2002).

Interests. Interests are also related to motivation in the sense that a person's interests affect his or her goals and level of attention (Hidi & Harackiewicz, 2000). Interests have both cognitive and affective components. The cognitive component refers to knowledge acquisition or cognitive processing that occurs when students are engaged in a topic, whereas the affective component refers to the positive associations people tend to feel for topics they are interested in (Hidi & Renninger, 2006). Researchers distinguish between two types of interest: individual and situational. Individual interest is a relatively stable trait developed over time with respect to a particular topic or subject. Situational interest, on the other hand, is more immediate and temporary and reflects certain external factors, such as characteristics of a given task. A student with individual interest in a subject will persist at the task longer than a student with merely situational interest. However, it may be possible to "hook" students by triggering situational interest in some aspect of the task (e.g., offering a degree of student choice or presenting the task as a collaborative group activity). This situational interest can sometimes blossom over time into a more lasting individual interest (Hidi & Harackiewicz, 2000).

Goals. A person's goals are related to his or her reasons for engaging with a task (Broussard & Garrison, 2004). Mastery goals focus on learning for the sake of learning, whereas performance goals are concerned with excelling in relation to others (Ames, 1992; Pintrich, 2003). Students holding mastery goals are more likely than those holding performance goals to have high self-efficacy, to believe success and failure are within their control, to persist at challenging tasks, and to use cognitive strategies associated with self-regulated learning. Such strategies include setting achievable goals, monitoring one's performance, evaluating progress, and framing success and failure with respect to effort and ability (Ames, 1992; Broussard & Garrison, 2004). As Eccles & Wigfield (2002) observe, mastery goals are associated with the strongest empirical evidence to date and have been linked to self-competence, self-concept, effort attributions, increased persistence at difficult tasks, and use of cognitive strategies related to monitoring, problem-solving, deep processing of information, and self-regulation.

In addition to the distinction between performance and mastery goals, there may also be a distinction between different types of performance goals. Performance approach goals are those that lead a student to tackle challenging academic tasks out of a desire to excel in relation to others, whereas performance avoid goals are those that cause a student to avoid challenging tasks so as not to risk appearing incompetent (Elliot & Church, 1997). Research focused on these different types of performance goals suggests that performance approach goals may not be as maladaptive in terms of student performance as once believed.

A wealth of empirical evidence on motivation exists, including research substantiating basic characteristics of the trait, such as domain specificity. Domain specificity is most likely due to individual differences in interests and self-concept across domains. In general, research suggests that the domain specificity of motivation and self-concept tends to increase with age, particularly as students accrue more educational experiences and as the curriculum begins to reflect departmentalization of academic subjects (Gottfried, et al., 2001).

Metacognition. John Flavell originally coined the term metacognition in the late 1970s to mean “cognition about cognitive phenomena,” or more simply “thinking about thinking” (Flavell, 1979, p. 906). Subsequent development and use of the term have remained relatively faithful to this original meaning. For example, researchers working in the field of cognitive psychology have offered the following definitions:

- “The knowledge and control children have over their own thinking and learning activities” (Cross & Paris, 1988, p. 131).
- “Awareness and management of one’s own thought” (Kuhn & Dean, 2004, p. 270).
- “The monitoring and control of thought” (Martinez, 2006, p. 696).

In cognitive psychology, metacognition is often defined as a form of executive control involving monitoring and self-regulation (Kuhn & Dean, 2004; McLeod, 1997; Schneider & Lockl, 2002). Schraw (1998) describes metacognition as a multidimensional set of general, rather than domain-specific, skills. These skills are empirically distinct from general intelligence, and may even help to compensate for deficits in general intelligence and/or prior knowledge on a subject during problem solving.

Constituent elements of metacognition. Metacognition has two constituent parts: knowledge about cognition and monitoring of cognition (Cross & Paris, 1988; Flavell, 1979; Paris & Winograd, 1990; Schraw & Moshman, 1995; Schraw et al., 2006; Whitebread et al., 1990). Several frameworks have been developed for categorizing types of cognitive knowledge and regulation. Table 2 organizes components from each of these frameworks to facilitate comparisons among them. For example, Flavell (1979) defines cognitive knowledge as knowledge about one's own cognitive strengths and limitations, including the factors (both internal and external) that may interact to affect cognition. He classifies such knowledge into three types: (1) "person" knowledge, which includes anything one believes about the nature of human beings as cognitive processors; (2) "task" knowledge, which includes knowledge about the demands of different tasks; and (3) "strategy" knowledge, which is knowledge about the types of strategies likely to be most useful.

Table 2. Typology of Metacognitive Components

Metacognitive Component	Type	Terminology	Citation
Cognitive knowledge	Knowledge about oneself as a learner and factors affecting cognition	Person and task knowledge	Flavell, 1979
		Self-appraisal	Paris & Winograd, 1990
		Epistemological understanding	Kuhn & Dean, 2004
		Declarative knowledge	Cross & Paris, 1988 Schraw et al., 2006 Schraw & Moshman, 1995
	Awareness and management of cognition, including knowledge about strategies	Procedural knowledge	Cross & Paris, 1988 Kuhn & Dean, 2004 Schraw et al., 2006
		Strategy knowledge	Flavell, 1979
	Knowledge about why and when to use a given strategy	Conditional knowledge	Schraw et al., 2006
Cognitive regulation	Identification and selection of appropriate strategies and allocation of resources	Planning	Cross & Paris, 1988 Paris & Winograd, 1990 Schraw et al., 2006 Schraw & Moshman, 1995 Whitebread et al., 2009
	Attending to and being aware of comprehension and task performance	Monitoring or regulating	Cross & Paris, 1988 Paris & Winograd, 1990 Schraw et al., 2006 Schraw & Moshman, 1995 Whitebread et al., 2009
		Cognitive experiences	Flavell, 1979
	Assessing the processes and products of one's learning, and revisiting and revising learning goals	Evaluating	Cross & Paris, 1988 Paris & Winograd, 1990 Schraw et al., 2006 Schraw & Moshman, 1995 Whitebread et al., 2009

Subsequent metacognition researchers have offered a slightly different framework for categorizing cognitive knowledge. For example, several researchers have used the concepts of declarative and procedural knowledge to distinguish cognitive knowledge types (Cross & Paris, 1988; Kuhn, 2000; Schraw et al., 2006; Schraw & Moshman, 1995). Kuhn and Dean (2004) characterize declarative cognitive knowledge broadly as epistemological understanding, or the student's understanding of thinking and knowing in general. Schraw et al. (2006) portray declarative cognitive knowledge as knowledge about oneself as a learner and what factors might influence one's performance. Paris and Winograd (1990) discuss the process of self-appraisal as reflection about personal knowledge states to answer the question, "Do I know this?" Finally, Cross and Paris (1988) define declarative cognitive knowledge specifically within the context of reading as awareness of the factors that might affect reading ability. On the other hand, procedural knowledge involves awareness and management of cognition, including knowledge about strategies (Cross & Paris, 1988; Kuhn & Dean, 2004; Schraw et al., 2006). Schraw et al. (2006) also distinguish conditional cognitive knowledge, which is knowledge of why and when to use a given strategy.

The other component of metacognition is monitoring of one's cognition, which many researchers have argued includes activities of planning, monitoring or regulating, and evaluating (Cross & Paris, 1988; Paris & Winograd, 1990; Schraw & Moshman, 1995; Schraw et al., 2006; Whitebread et al., 2009). Planning involves identification and selection of appropriate strategies and allocation of resources, and can include goal setting, activating background knowledge, and budgeting time. Monitoring or regulating involves attending to and being aware of comprehension and task performance and can include self-testing. Finally, evaluation is defined as “appraising the products and regulatory processes of one’s learning,” and includes revisiting and revising one’s goals (Schraw et al., 2006, p. 114).

Researchers have observed a relationship between cognitive knowledge and cognitive monitoring. For example, Flavell (1979) argues that metacognitive experiences that allow one to monitor and regulate one’s cognition play a major role in the development and refinement of metacognitive knowledge. In turn, Schraw (1998) cites a number of empirical studies demonstrating that cognitive knowledge appears to facilitate cognitive regulation. He notes that such studies have found cognitive knowledge and cognitive regulation to be correlated with one another at about $r = .50$.

How Are These Constructs Related To One Another?

Research on these skills suggests they are interrelated in complex ways. For example, critical thinking and creativity are often expressed together. Flexibility, a subskill of creativity, supports critical thinking to the extent that being open-minded and willing to see from diverse perspectives is a key disposition when making decisions or solving problems. Likewise, the critical thinking skills of analysis, synthesis, and evaluation are often identified as key components of the creative process (Paul & Elder, 2006). Interplay between creative and critical thinking skills is essential during creative problem solving. When presented with problems, contradictions, or challenges, the creative thinker generates many possible solutions in a search for meaningful new connections. Critical thinking is then required in order to analyze and evaluate the possible solutions (Treffinger et al., 2002). Critical thinking skills are also necessary in the overall creative process of generating a new idea or product. To be creative, an idea or product must be both novel and have value. Each potential idea must be analyzed and evaluated to determine whether it has value and is worth pursuing (Sternberg, 2006a).

Metacognition and critical thinking are also reciprocally related. Flavell (1979) and Martinez (2006) maintain that critical thinking is actually subsumed under metacognition. For example, Martinez defines critical thinking as “evaluating ideas for their quality, especially judging whether or not they make sense,” and sees it as one of three types of metacognition, along with metamemory and problem solving (p. 697). Schraw et al. (2006), however, see both metacognition and critical thinking as being subsumed under self-regulated learning, which they define as “our ability to understand and control our learning environments” (p. 111). Self-regulated learning entails metacognition, motivation, and cognition, which includes critical thinking (Schraw et al., 2006). At the very least, metacognition can be seen as a supporting condition for critical thinking, to the extent that monitoring the quality of one’s thought makes it more likely that one will engage in high-quality (critical) thinking. In turn, the ability to critically evaluate one’s own arguments and reasoning is necessary for self-regulated learning.

Critical thinking may also support and be supported by motivation. The disposition to think critically has been defined as the “consistent internal motivation to engage problems and make decisions by using critical thinking” (Facione, 2000, p. 65). Thus, student motivation is viewed as a necessary precondition for the exercise of critical thinking skills and abilities. Similarly, Halonen (1995) notes that a person’s propensity or disposition to demonstrate higher-order thinking relates to his or her motivation. Halpern (1998) argues that effort and persistence are two of the principle dispositions that support critical thinking, and Paul (1992) maintains that perseverance is one of the “traits of mind” that render someone a critical thinker. Thus, motivation appears to be a supporting condition for critical thinking in that unmotivated individuals are unlikely to exhibit critical thought.

On the other hand, a few motivation researchers have suggested the relationship goes the other way. In particular, motivation research suggests that difficult or challenging tasks, particularly those emphasizing higher-order thinking skills, may be more motivating to students than easy tasks that can be solved through rote application of a predetermined algorithm (Turner, 1995). Pintrich's framework holds that cognition and motivation affect one another, both affect academic achievement, and that both, in turn, are affected by the social context of learning (Linnenbrink & Pintrich, 2002; Pintrich, 2003).

Motivation and metacognition are also reciprocally related. Both concepts are subsumed under the construct of self-regulated learning, which "emphasizes autonomy and control by the individual who monitors, directs, and regulates actions toward goals of information acquisition, expanding expertise, and self improvement" (Paris & Paris, 2001, p. 89). Metacognition entails the management of affective and motivational states, and metacognitive strategies can improve persistence at challenging tasks (Cross & Paris, 1988; Martinez, 2006). As Turner (1995) observes, "because strategy use is effortful and time-consuming and because it requires active monitoring and evaluation, it is an indicator of students' cognitive engagement" (p. 419). Effortful control, which refers to the ability to monitor and regulate the impact of emotions and motivational states on one's performance, is one aspect of the executive functioning inherent in metacognition. Research suggests that effortful control among preschool- and elementary-age children is associated with better social relationships at school, higher academic engagement, and improved achievement (Eisenberg, 2010).

Task oriented, intrinsic motivation has also been identified as an essential component of creativity. Studies by Amabile suggest that people who choose to be creative are driven by a love for what they are doing rather than by any potential rewards the creative endeavor may bring them (Sternberg, 2010). In fact, truly creative ideas often meet resistance from those who view new ideas (and the people behind them) with suspicion and/or have a vested interest in maintaining the status quo (Sternberg, 2010). One must be motivated to be creative despite the strong potential for push-back from society (Sternberg, 2006a). Motivation is also linked to creativity through another essential affective component—intellectual risk taking. A person's motivational goals affect the propensity to take intellectual risks, which can subsequently affect that person's decision to engage in creative endeavors. People who hold learning or mastery goals tend to view intellectual risks as opportunities to receive feedback that they can use to grow and improve. Such people are more likely to take intellectual risks than are people who hold performance or ego-oriented goals, for whom it is important to excel in relation to others. People holding the latter type of goals are more likely to avoid taking intellectual risks, and less likely to be creative, out of a fear of failure (Clifford, 1991).

Finally, several of these skills are related to collaboration. Research suggests that providing students with opportunities to work together may stimulate students' critical thinking, motivation, and metacognition. Collaborative learning approaches are believed to encourage development of critical thinking to the extent that working with others stimulates cognitive conflict, which occurs when a student encounters new information or experiences that conflict with his or her existing frameworks (Dillenbourg et al., 1996). Similarly, collaborative approaches promote metacognitive development among students when they are prompted to provide elaborated explanations and make their thinking and reasoning visible. Providing such explanations can help students become more aware of their own thinking (Dekker et al., 2006; Kramarski & Mevarech, 2003; Kuhn & Dean, 2004; Van Boxtel et al., 2000). Peer encouragement may improve task engagement, and the novelty of collaborative learning tasks causes students to pay closer attention (Bossert, 1988). Working with others is a way of enhancing interest in the task (Hidi & Harackiewicz, 2000). Collaboration also provides opportunities for peer modeling, and students may find models of successful student performance to be more motivating than models of teacher performance (Webb et al., 2001).

How Do Researchers Measure These Constructs?

Researchers have approached assessment of each of these skills in a variety of ways, although there are areas of overlap in the approaches used

Self-reports. It is quite common for researchers to administer self-report surveys or inventories in an attempt to capture examinee skills, attitudes, and dispositions. For instance, creativity researchers use self-reports to examine personality traits and attitudes, such as risk taking, openness to new experiences, and motivation (Lubart & Guignard, 2004). Self-report tools also allow individuals to rate their own creativity-related skills, achievement, behaviors and abilities and include biographical inventories, questionnaires, and personal checklists (Lubart & Guignard, 2004; Treffinger et al., 2002). Similarly, motivation is frequently assessed via self-report measures (Broussard & Garrison, 2004; Elliott & Church, 1997; Gottfried, 1990; Guay et al., 2010; Guthrie et al., 2000; Lange & Adler, 1997; Miller & Meece, 1997; Ryan et al., 1990). Such instruments usually include questions organized under several subscales, including interest, attributions, self-perception and self-efficacy, preference for challenge, curiosity, mastery orientation, persistence, and enjoyment of learning. Self-reports have also been used to measure metacognitive skills. For instance, researchers have administered self-report measures of students' use of strategies, perceived helpfulness of strategies, awareness of one's own performance, and ability to plan and monitor progress towards goals (Cross & Paris, 1988; Kramarski & Mevarech, 2003; Pintrich & DeGroot, 1990; Sperling et al., 2002).

Although self-reports may be relatively easy and cost-effective to administer, their limitations are well-documented. Within creativity, for example, self-reports are limited with respect to the "completeness and accuracy of any self-description, the comparability of data across settings, the stability of self assessments over time, or the correlation between self-ratings and other external criteria of creativity" (Treffinger et al., 2002, p. 42). Whitebread et al. (2009) argue that self-report methods, such as the use of rating scales or questionnaires that ask respondents to describe their use of particular strategies, rely too heavily on verbal ability.

Turner (1995) argues that such measures may be particularly inappropriate for children. Research suggests that children often have difficulty providing the type of generalized response that is commonly sought in self-report instruments. Children tend to instead interpret just-experienced events rather than summarize across a range of situations and content areas. In addition, children have a tendency toward positive response bias because they are more inclined than older students to be optimistic. As Turner (1995) explains, children have difficulty separating their efforts and intentions from their actual behavior. Thus, children who intend to exert a lot of effort may mistakenly believe that they have actually done so. Children are also highly susceptible to social desirability, another source of positive response bias. When using these instruments with early elementary-age children, researchers typically have to modify instruments, such as reducing the language load, simplifying rating scales, and reading items aloud to students (Gottfried, 1990; Guay et al., 2010; Miller & Meece, 1997).

Global Rating Scales. Another popular method for capturing these skills is through the use of a global rating system completed by parents or educators. For instance, creativity rating scales ask respondents to rate other people's creative skills, attributes or abilities based on specific descriptions of creative traits (Treffinger et al., 2002). Global rating scales are also used to measure motivation. For example, parents and teachers have been asked to rate traits such as students' mastery-oriented learning behaviors and intrinsic versus extrinsic goal orientation (Gottfried, 1990; Lange & Adler, 1997). The Personal Potential Index, developed by ETS, is a global rating system used by graduate admissions programs to better discriminate amongst candidates for graduate school admission. This is a standardized instrument that asks faculty members familiar with the candidate to rate him or her on six dimensions: knowledge and creativity, communication skills, teamwork, resilience, planning and organization, and ethics and integrity (Kyllonen, 2008).

As Treffinger et al. (2002) note, many factors influence the accuracy of global ratings, including the number and types of opportunities for raters to observe the subjects, the willingness of raters to focus scores only on the criteria provided in the assessment tool, and the raters' overall understanding of the behaviors to be rated. Similarly, Kyllonen (2008) notes that although such ratings systems are less susceptible to coaching or faking than self-report measures, ratings are prone to the same type of subjectivity that plagues course grades and traditional recommendation letters. In addition, such ratings are susceptible to halo effects, particularly when parents or teachers are assigning ratings along a number of dimensions or across a number of occasions (Zhuang et al., 2008). Kyllonen (2008) points out a number of potential ratings adjustments that could be made to help correct for rater bias, including leniency or severity. However, accurate adjustments of this sort become possible only after a sufficient pool of ratings becomes available that allows (1) the same candidates to be rated by multiple raters and (2) individual raters to accumulate ratings for many candidates over time.

Standardized Assessments. Several of these skills have been assessed using standardized assessments employing either traditional multiple-choice items or more open-ended prompts. For example, existing published assessments of critical thinking are numerous, and include the California Critical Thinking Skills Test (Facione, 1990), the Cornell Critical Thinking Tests (Ennis & Millman, 2005), and the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980). Such assessments utilize multiple-choice items designed to assess component critical thinking skills, such as deductive reasoning, inductive reasoning, drawing conclusions, evaluating arguments, and so forth.

There are also tests of creativity that employ multiple-choice items. For instance, Sternberg (2006a) designed a series of tests that target a person's ability to switch between conventional and unconventional thinking, a trait that he has identified as correlating with creative ability. The tests consist of reasoning problems for which there is a single best answer. However, the problems are based on either novel or counterfactual premises. Test takers are asked to work through the reasoning problems as though the hypothetical or counterfactual situation were true (Sternberg, 2006a). Finally, situational judgment tests (SJTs) have been widely used in personnel selection contexts to assess examinee judgment in work-related situations. Typically, the examinee reads a detailed description of a scenario and is asked to select the most appropriate response or to indicate what he or she would do in that situation. Such measures have been used to capture communication, teamwork, and collaboration skills, attitudes, and dispositions (McDaniel et al., 2001; Zhuang et al., 2008).

The benefit of machine-scorable approaches is that they can be administered easily and cost-effectively. Moreover, to the extent that multiple-choice items make more efficient use of testing time, the domain can be more sampled more broadly than can be accomplished with constructed response items. However, methods relying exclusively on multiple-choice type items tend to assess only limited aspects of what it means to be a critical thinker or a creative person (Norris, 1989; Sternberg, 2006a). Because they use scenarios that tend to be drawn from real experience, SJTs may offer a more authentic context in which to elicit examinee skills. However, researchers also note that SJT approaches are susceptible to faking or coaching and examinee performance on these types of items can be confounded with verbal ability when the scenarios require large amounts of reading. In addition, it can be difficult to determine a single best response, requiring the input of experts. For domains such as collaboration, it is not clear what constitutes expertise (Zhuang et al., 2008).

There are also more open-ended forms of standardized assessments designed to tap 21st century skills. For example, the Ennis-Weir Critical Thinking Test (Ennis & Weir, 1985) is a standardized test that uses essay questions to assess students' ability to read and evaluate arguments. The Multistate Essay Exam and the Multistate Performance Test (both standardized components of the Bar Exam used to license lawyers) are assessments of examinee reasoning and argumentation skills that require candidates to respond in writing to hypothetical legal scenarios, discern relevant from irrelevant case facts, and analyze legal materials and issues (Case, 2011). The Ennis-Weir is a domain-general critical thinking measure, whereas the components of the Bar Exam are clearly anchored in the legal domain.

The Torrance Tests of Creativity, the most widely used creativity assessment, is a constructed response test with two versions, one verbal and one figural. The verbal battery targets five types of activities: ask-and-guess, product improvement, unusual uses, unusual questions, and "just suppose." For each item, the examinee responds in writing to a stimulus, which includes a picture. The figural battery requires examinees to manipulate and add to incomplete figures or drawings to create a new object or picture (Kim, 2006).

Tests of divergent thinking usually require the participant to generate as many possible responses to a stimulus (uses of objects, ways to complete a picture, titles to a story, etc.) within a set time limit and are typically judged for fluency (number of unique responses), flexibility (variety of types of responses), and originality (responses that show evidence of thinking outside the box). Elaboration (amount of details within responses) is also sometimes scored (Lubart & Georgsdottir, 2004).

These open-ended approaches to measuring 21st century skills may be able to tap knowledge and skills at deeper levels and in more authentic contexts than traditional item formats. However, one significant limitation of these approaches is the cost associated with scoring. All of the open-ended assessment examples identified above must be hand-scored, which requires extensive training and monitoring if it is to be done consistently. Such requirements have obvious implications for the ability to administer and score open-ended approaches on a large-scale in an efficient and cost-effective manner.

Observational Measures. Finally, researchers have used a number of different observational techniques for cataloging student behaviors relevant to 21st century skills. These techniques differ from one another primarily with respect to whether the context of the measure is classroom- or laboratory-based. In other words, such measures differ in relation to whether they are administered during the course of regular school work occurring in the classroom or whether they are administered in a lab using non-school-related tasks.

Metacognition and motivation are skills not traditionally assessed in school as part of the regular curriculum. Thus, these skills are often studied in a lab environment that is somewhat artificial or contrived, in the sense that it is not connected to school learning. Moreover, these studies often focus on isolated aspects of the whole construct. For example, some metacognition studies focus narrowly on metamemory. Flavell (1979) describes assessment tasks that asked children to study a set of items until they were sure they could remember them completely. Children were then tested on their ability to recall all the items. Schneider (2008) observes that the most studied type of procedural metamemory is self-monitoring. Assessments designed to capture this ability include ease of learning judgments, judgments of learning, and feelings of knowing. Ease of learning judgments typically ask students to study a set of test materials for a short amount of time and then assess their abilities to remember the material. After the students are tested on the material, their performances are compared to their initial predictions. Feeling of knowing judgments ask subjects to identify by name a series of pictures; when subjects cannot recall the word for a particular picture, they are asked whether they would be able to identify the word if it were shown to them. These predictions are then compared to their actual abilities to recognize the correct term among a list of options.

Other metacognition researchers have focused on subjects' theory of mind using location-false belief, contents-false belief, deceptive pointing, and appearance-reality tasks (Carlson & Moses, 2003). Each of these tasks involves cognitive conflict in some way, in the sense that successful performance requires subjects to suppress impulsive responses and to produce a response that is incompatible with the dominant response. For example, in one standard contents-false belief task, children are shown a common, brand-name box of bandages and asked to predict what is inside. The box is then opened and children are shown that it actually contains crayons. Another investigator then enters the room and is shown the closed box. Children are asked to speculate about what the second experimenter believes is in the box. A standard appearance-reality task attempts to train children to respond “day” when shown a picture of the moon and “night” when shown a picture of the sun.

Similarly, most of the empirical studies included in the Deci et al. (1999) meta-analysis on motivation focus on subjects’ free-choice persistence, which is typically a measure of the amount of time spent on a laboratory activity (e.g., puzzles, tangrams, etc.) once extrinsic reward conditions have been suspended.

On the other hand, there are also examples of classroom-based measures of motivation and metacognition. For instance, Turner (1995) constructed a behavioral measure of motivation that was administered during the course of regular reading instruction in the classroom. This measure included aspects related to effective strategy use, persistence, and volitional acts. Turner defines strategies as “intentional, deliberate actions that learners invoke to solve a specific problem or meet a particular goal” (p. 419). Effective strategy use behaviors include the use of general strategies, such as rehearsal, elaboration, and organization, and task-specific strategies, such as decoding and comprehension during reading. Behavioral indicators of persistence include asking for help, asking oneself questions, or talking oneself through a task.

Hennessey (1999) developed six categories to characterize the various levels of metacognition evident in students’ discourse as they constructed or revised representations of their science conceptions in small groups during regular science instruction. Whitebread et al.

(2009) developed an observational checklist with 22 items to measure metacognition and self-regulation in children between the ages of 3 and 5. This checklist identifies a range of student behaviors—both verbal and nonverbal—theorized to represent metacognitive knowledge, metacognitive regulation, and emotional and motivational regulation. Teachers are to code metacognitive events observed during the course of individual or group learning by rating individual students on each behavior.

Researchers studying collaboration have also used a variety of observational tools to capture students' interactions while they grapple with academic problems. Dillenbourg et al. (1996) base their collaboration framework on conversation models developed in the field of linguistics. The authors are primarily interested in student negotiation, which they define as a process by which students attempt to attain agreement on aspects of the task or interaction. They identify four types of negotiation behaviors that can be observed during interaction: mutual adjustment occurs when partners refine their positions; competitive argumentation occurs when one partner tries to convince the other to adopt his position; "standing pat" is when one student uses another as a resource by eliciting a proposal; and negotiation of meaning is evidenced by "repair sequences," in which misunderstandings become evident and are explicit targets of discussion.

Roschelle and Teasley (1995) have delineated a number of conversational strategies for achieving shared understanding. These strategies are said to indicate deep levels of collaborative interaction and include taking turns, socially distributed productions, repairs, narrations, and nonverbal actions or gestures. Mercer (1996) recorded around 60 hours of classroom talk with 50 children between the ages of 5 and 13 while they worked in small groups on collaborative tasks. He categorized student talk into three types: disputational talk, cumulative talk, and exploratory talk. Disputational talk is characterized by disagreement and individualized decision-making, with few attempts to pool resources, or to offer or accept constructive criticism. In cumulative talk, speakers build positively but uncritically on what the other has said. Finally, exploratory talk is when partners engage critically but constructively with one another. Students engaging in such talk offer statements and suggestions for joint consideration. These may be challenged, but all arguments are justified and reasons are provided.

Finally, Webb et al. (1998) code contributions of individual students to group discussion according to their cognitive level. Thus, high-level participation includes making or defending suggestions for how to answer a particular item, asking questions about a suggestion, or paraphrasing a suggestion. Medium-level participation includes copying someone else's suggestion, repeating it verbatim, or agreeing with what was said without further elaboration. Low-level participation entails listening or watching without making any substantive contribution or inquiry. Finally, students who manipulate materials or write answers without referencing other group members are coded as working alone.

The benefit of such observational techniques is that they are based on students' actual behaviors, both verbal and nonverbal, as they engage with tasks. Moreover, observational tools designed for use with regular academic tasks (as opposed to laboratory tasks) have the added benefit of being more related to in-school learning, which has two implications: 1) students may find the tasks more personally meaningful and 2) they are closer to the types of problems students will face in future educational and employment settings. One obvious limitation of these techniques is their heavy reliance on close observation by a professional educator or researcher to observe, record, and interpret student behaviors. Such labor intensity suggests these observational approaches may not be feasible for large-scale testing.

Recommendations and Discussion

Given the answers to our research questions, we propose several recommendations related to assessing 21st century skills.

Assessment systems should provide multiple measures that support triangulation of inferences. A number of challenges in measuring 21st century skills suggest the need for multiple measures that either 1) represent multiple assessment modes or 2) sample from multiple content domains to permit triangulation of inferences. First, the limitations of any single measures (e.g., limited validity for self-reports) are well-documented and were described in the previous section. Second, a number of constructs suffer definitional ambiguity, with disagreement even among experts. As a result, different measures tap different aspects of the same construct (Treffinger et al., 2002). For example, as noted by Runco (2003) approaches to defining and assessing creativity tend to differ with respect to whether they focus on aspects of the creative process (such as divergent thinking or ideation) or creative products and outputs (such as works of art or creative performances). Indeed, Han and Marvin (2002) found that measures related to the creative process (such as divergent thinking) were not highly correlated with actual creative performances.

Similarly, assessments of collaboration tend to focus on either collaborative processes (such as communication, negotiation, and compromise) or the quality of collaborative products. As Webb (1995) notes, assessment developers need to be clear about whether the purpose of the assessment is to capture process or product, because group interactions that are considered maladaptive for collaborative processes, such as free riding or social loafing, can actually be effective for groups if the goal is to maximize productivity.

A third reason to use multiple measures is definitional complexity. Several of these skills (e.g., creativity, metacognition, motivation) entail multiple sub-components. For example, metacognition is a complex construct, involving cognitive knowledge and cognitive regulation. Moreover, there are multiple types of cognitive knowledge (declarative, procedural, conditional) as well as different types of cognitive regulation (planning, monitoring or regulating, and evaluating). Moreover, each of the skills reviewed in this paper entails both cognitive and noncognitive or affective components. Cognitive components of these constructs include knowledge and strategies, whereas noncognitive components include attitudes, personality traits, and dispositions. Cognitive and noncognitive elements are commonly confounded in practice. To provide a more complete picture of student skills, both components should be addressed, ideally using complementary assessment modes (Ku, 2009). To assess metacognition, for example, one might pair a standardized assessment of students' declarative, procedural, and conditional knowledge about strategies with a student self-rating scale designed to capture more affective components, such as regulation of emotion and motivation during task engagement.

Finally, research suggests that several of these skills (e.g., critical thinking, creativity, and motivation) exhibit some degree of domain specificity, which implies that a student may demonstrate critical thinking in one domain (e.g., science), but fail to do so in another (e.g., math). For creativity and critical thinking, domain specificity stems from content-specific knowledge required to exercise the skill, which suggests that measurement of the skill may be confounded in practice with measurement of related content knowledge. Thus, assessments should sample from multiple domains. Student performance in multiple domains could then be expressed as a profile reflecting relative strengths and weaknesses.

Assessment tasks should be of sufficient complexity and/or offer sufficient challenge. Research suggests that the challenge level or complexity of assessment tasks is a key consideration in accurately eliciting and capturing 21st century skills. For example, task difficulty affects students' engagement, with most researchers arguing that difficulty or challenge level impacts motivation through students' sense of competence (Ames, 1992). That is, tasks that are perceived as too easy can diminish students' engagement because completion of the task does not promote a sense of competence (Lange & Adler, 1997). On the other hand, tasks that are perceived as too challenging may prompt challenge-avoidance behaviors, such as disengagement and low effort, if students are not confident that they can excel at the task (Covington, 1992, as cited in Stipek, 1996). Thus, tasks of moderate difficulty are argued to elicit increased student persistence, more varied strategy usage, and greater task interest (Turner, 1995).

Complex tasks are similarly important for assessing collaboration. For example, Dillenbourg (1999) points out that trivial, obvious, and unambiguous tasks provide few opportunities to observe student negotiation because there is nothing about which to disagree. Moreover, misunderstandings may actually be important from a learning standpoint, as they force participants to construct explanations, give reasons, and justify their positions. Metacognition researchers have also highlighted the importance of providing complex tasks that allow multiple representations of concepts and afford students opportunities to identify and resolve conceptual conflict (Kramarski & Mevarech, 2003). Finally, Fischer et al. (2009) demonstrated that tasks relying on stimulus materials designed to evoke cognitive conflict (i.e., that reflected uncertainty, ambiguity, disorganization, and contradiction) better elicited critical thinking skills than tasks that used stimulus materials that were orderly, well-organized, and coherent.

Assessments should include open-ended and/or ill-structured tasks. Researchers agree that traditional approaches to standardized assessments that rely heavily on forced-choice item formats can assess only limited aspects of these constructs (Norris, 1989; Ku, 2009; Sternberg, 2006a). Rather, researchers recommend the use of open-ended problems (Ku, 2009; Sternberg, 2006a; Turner, 1995). According to Turner (1995), "open" tasks allow students to decide what relevant information to use or how to use the information to solve the problem. These types of tasks require more metacognition and decision-making. In contrast, "closed" tasks are characterized by more teacher control and structure. With closed tasks, teachers indicate both the information to be used and what the expected solution will look like, and tasks typically emphasize a single, correct solution. In the context of motivation research, "open" tasks may also imply a higher level of examinee autonomy. In other words, tasks that are differentiable by student interest or ability or that permit students to make choices are more likely to elicit motivated behaviors (Ames, 1992; Stipek, 1996; Turner, 1995).

Moreover, researchers recommend using so-called "ill-structured" problems to assess critical thinking, motivation, and collaboration (Moss & Koziol, 1991; Turner, 1995; Webb, Nemer, Chizhik, & Sugrue, 1998). Ill-structured problems are those with no clearly defined parameters, no clear solution strategies, and either more than one correct solution, or multiple ways of arriving at an acceptable solution. For example, critical thinking prompts should require students to go beyond the available information in the task to draw inferences or make evaluations. In addition, problems should have more than one plausible or defensible solution, and there should be sufficient information and evidence within the task to enable students to support multiple views (Moss & Koziol, 1991).

Within collaboration, researchers emphasize the importance of ill-structured tasks that cannot be solved by a single, competent group member (Webb et al., 1998). Such tasks require knowledge, information, skills, and strategies that no single individual is likely to possess. For example, Salomon and Globerson (1989) define additive tasks as those “where performance depends on the maximal contribution of all members,” giving the example of a tug- of-war game. Both additive tasks and ill-structured tasks are more likely to encourage full group participation to the extent that the task cannot be completed by a single, competent person. In fact, Webb et al. (1998) observe that when ill-structured tasks are used, all group members are more likely to participate actively, even in groups featuring a range of student ability.

Assessments should use tasks that establish meaningful and/or authentic, real-world problem contexts. Researchers working in the fields of creativity, motivation, and critical thinking have emphasized the importance of assessment tasks that provide authentic problem contexts that students will find personally meaningful (Ames, 1992; Halpern, 1998; Hidi & Harackiewicz, 2000; Okuda et al., 1991; Runco, 1993; Turner, 1995). For example, Turner (1995) points out that authentic tasks tend to be more extended, more student-directed or individualized, require integration of multiple skills, and “make deliberate use of real-world social and physical contexts” (p. 416). Similarly, Halpern (1998) argues that authentic problem contexts should be based on simulations that approximate real-world problems and issues. Okuda et al. (1991) compared typical divergent thinking tasks with those that utilized real-world problem contexts—situations students would likely face in school—and found that student performance on tasks using real-world contexts correlated with actual creative performances occurring in a natural environment at much higher levels than did tasks using artificial or contrived problem contexts.

Assessment tasks should strive to make student reasoning and thinking visible. For critical thinking, metacognition, and collaboration, tasks should make student thinking and reasoning visible by requiring students to provide evidence or logical arguments in support of judgments, choices, claims, or assertions. For example, Norris (1989) argues that testing validly for critical thinking requires that we observe an examinee's process of thinking. Similarly, to evaluate an examinee's metacognitive abilities, we must first have access to his or her thinking. Less obviously, our ability to observe students' collaboration skills is dependent on them being able to share their own thinking and reasoning with teammates via verbal or nonverbal behaviors (Mercer, 1996; Van Boxtel, 2000; Webb, 1991). Thus, tasks should be structured in ways that encourage or require students to share their thinking and reasoning, either with one another or with the examiner.

One common approach used in the metacognition research is to employ verbal protocols or "think alouds, where examinees reflect out loud on their performance as they engage with a task." Clearly, such approaches are too time-intensive and costly to implement on a large scale. However, this point reinforces the importance of open-ended tasks that offer opportunities for students to explain their thinking. One recommendation for making student thinking visible in the context of a multiple-choice test is to require students to provide a rationale or justification for each choice (Kennedy et al., 1991; Ku, 2009; Norris, 1989). This latter approach is somewhat evident in PARCC's specifications for evidence-based selected response items, which are to be used for assessing reading skills on the mid-year, performance-based, and end-of-year components of its comprehensive assessment system. In responding to these items, students are required to flag particular portions of the passage or stimulus that form the basis for their selected response.

Assessments should explore innovative approaches to address scalability concerns. Collectively, the research on assessing 21st century skills raises a number of concerns regarding the feasibility of implementing such assessments at scale in an efficient and cost-effective manner. However, the research also suggests that innovative approaches to administration and scoring that rely on new technologies and non-traditional psychometric models may constitute viable means of overcoming these obstacles.

As previously discussed, one of the biggest limitations of current approaches to measuring these skills is their heavy reliance on examiner observation, which is time-consuming and costly to administer. Technology-enhanced item formats, such as simulations, tasks that use multimedia stimuli, or items that allow examinees to record their responses using innovative media present one way of addressing this challenge. Technology-enhanced items (TEIs) may be administered online, which allows assessment developers to take advantage of new technologies that can supplement human observation, such as audio- or video-recording devices, chat software, programs like Power Point, and basic word processing programs (Dolan et al., 2011). For example, more recent approaches to assessing collaboration, known as computer-supported collaborative learning, use technology tools, such as cell phones and networked computers, to both mediate and record the results of collaboration. Students can work with one another or with automated “agents” in a synchronous or asynchronous fashion via networked computers, chat programs, and video- or teleconferencing capabilities. During interactions, comprehensive logs of student actions and exchanges are recorded for scoring or coding (Curtis and Lawson, 2001; Kreijns et al., 2003; Stahl et al., 2003). Portions of these exchanges or interactions—including more extended text-based exchanges or interaction sequences with stimuli—are potentially machine-scorable through artificial intelligence (AI).

Another limitation of current assessment approaches concerns the costs associated with scoring open-ended tasks, such as extended constructed response or complex performance-based tasks. Once again, innovative assessment approaches are beginning to address these concerns. First, automated scoring has advanced in recent years, suggesting that an increasing array of item types in several domains (not just ELA) can be reliably and validly scored via AI (Streeter, Bernstein, Foltz, & DeLand, 2011). Second, certain types of TEIs—particularly those classified by Scalise and Gifford as intermediate constraint items—are experienced as open-ended problems by examinees, but can be scored as quickly and reliably as more traditional selected response items (Scalise & Gifford, 2006). Thus, although TEIs are more expensive to develop than traditional items, cost efficiencies can be gained in scoring, particularly for very large numbers of examinees.

TEIs with authentic and rich problem contexts also have the added benefit of being able to tap aspects of these constructs that are not well-measured using more traditional item types (Bennett et al., 2003; Huff & Sireci, 2001; Mislevy et al., 2002). By offering unique and varied student response modes, TEIs may mitigate the impact of construct-irrelevant variance, such as limited writing skills on an assessment of creative thinking in math (Dolan et al., 2011). Finally, the novelty of TEIs may make them more engaging for students, thus increasing the likelihood of eliciting (and observing) student motivation (Huff & Sireci, 2001).

Finally, as described in a previous section, research on these constructs suggests they are inter-related in complex ways. Such connections suggest that a cost-efficient and valid strategy for assessing such skills on a large-scale may be to design tasks that require examinees to use multiple skills in concert. Such integrated tasks would, however, give rise to multidimensionality in student responses, which is not handled well in commonly-used psychometric models, such as unidimensional IRT approaches (Bennett et al., 2003). Thus, innovative approaches to modeling student responses that reflect the multidimensionality inherent in complex, integrated tasks should be considered, such as multidimensional scaling techniques.

Next Steps

Development of the next generation of assessment systems is already underway via the Race to the Top assessment consortia, whose plans indicate that some of these recommendations may be implemented at large scale within the next few years. For example, current plans suggest the use of complex, open-ended performance tasks that feature real-world scenarios focused on making student reasoning, thinking, and argumentation visible, through a combination of both traditional constructed-response items and TEIs. Additional research is needed on a number of topics to enable full implementation of these recommendations, including (1) current capabilities and limitations of automated scoring, (2) the use of data mining techniques to extract usable and informative data from immersive online environments, (3) the benefits and liabilities of TEIs, particularly with respect to construct (ir)relevance arising from accessibility and bias issues, (4) the feasibility and palatability of implementing multdimensional scaling and (5) whether assessments of this sort can support scores reliable enough to justify high-stakes decisions about students. Addressing these (and other) challenges in the coming years will require assessment developers to think in 21st century ways.

References

Ames, C. (1992). Classrooms: Goals, structure, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31*(3), 285–302.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148.

Beghetto, R. A. (2009). Correlates of intellectual risk taking in elementary school science. *Journal Research in Science Teaching, 46*(2), 210-223.

Beghetto, R. A. (2010). Creativity in the classroom. In J. Kaufman & R. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 447–463). New York, NY: Cambridge University Press.

Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (2003). *Assessing complex-problem solving performances* (ETS Report No. RM-03-03). Princeton, NJ: Educational Testing Service.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2010). Defining 21st century skills. Assessment and teaching of 21st century skills draft white paper. The University of Melbourne.

Blatchford, P., Kutnick, P., Baines, E., & Galton, M. (2003). Toward a social pedagogy of classroom group work. *International Journal of Educational Research, 39,* 153-172.

Bossert, S. T. (1988). Cooperative activities in the classroom. *Review of Research in Education, 15*(1988-1989), 225–250.

Broussard, S. C., & Garrison, M. E. B. (2004). The relationship between classroom motivation and academic achievement in elementary school-aged children. *Family and Consumer Sciences Research Journal*, *33*(2), 106–120.

Carlson, S. M. & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053.

Case, R. (2005). Moving critical thinking to the main stage. *Education Canada, 45*(2): 45–49.

Case, S. M. (2011). Assessment of critical thinking and problem solving on the multistate bar exam. Presentation delivered at the National Research Council's 21st Century Skills Workshop, Irvine, CA. Retrieved from http://www7.nationalacademies.org/bota/21st_Century_Workshop_Case.pdf.

Clifford, M. M. (1991). Risk Taking: Theoretical, Empirical, and Educational Considerations. *Educational Psychologist* 26, no. 3 and 4, 263-297.

Cross, D. R. & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80*(2), 131–142.

Curtis, D. D. & Lawson, M. J. (2001). Exploring collaborative online learning. *The Journal of Asynchronous Learning Networks*, *5*(1), 21-34.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Dekker, R., Elshout-Mohr, M., & Wood, T. (2006). How children regulate their own collaborative learning. *Educational Studies in Mathematics,* 62, 57–79.

Dillenbourg, P. (1999). What do you mean by 'collaborative learning?' In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp.1–19). Oxford: Elsevier.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.

Dolan, R. P., Goodman, J., Strain-Seymour, E., Adams, J., & Sethuraman, S. (2011). *Cognitive lab evaluation of innovative items in mathematics and English/ language arts assessment of elementary, middle, and high school students*. Research Report. Iowa City: Pearson. Retrieved from http://www.pearsonassessments.com/hai/images/tmrs/Cognitive_Lab_Evaluation_of_Innovative_Items.pdf.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Eisenberg, N. (2010). Self-regulation and school readiness. *Early Education and Development, 21*(5), 681-698.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218-232.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44–48.

Ennis, R. H., & Millman, J. (2005). *Cornell critical thinking test: Level X* (5th ed.). Seaside, CA.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: The California Academic Press.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relation to critical thinking skill. *Informal Logic, 20*(1), 61–84.

Fall, R. Webb, N., & Chudowsky, N. (1997). *Group discussion and large-scale language arts assessment: Effects on students' comprehension*. CSE Technical Report 445. Los Angeles, CRESST.

Fischer, S. C., Spiker, V. A., & Riedel, S. L. (2009). *Critical thinking training for army officers, volume 2: A model of critical thinking.* (Technical Report). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911.

Frederickson, N. & Ward, W. C. (1978). Measures for the study of creativity in scientific problem-solving. *Applied Psychological Measurement*, *2*(1), 1-24.

Gillies, R. M. (2003). The behaviors, interactions, and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology 95*(1), 137–147.

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology,* 98(4), 732–749.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82*(3), 525–538.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*(1), 3–13.

Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology, 80*(4), 711–735.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331–341.

Halonen, J. S. (1995). Demystifying critical thinking. *Teaching of Psychology, 22*(1), 75–81.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455.

Han, K. S. & Marvin, C. (2002). Multiple creativities? Investigating domain-specificity of creativity in young children. *Gifted Child Quarterly, 46*(2), 98-109.

Hennessey, M. G. (1999, March). Probing the dimensions of metacognition: Implications for conceptual change teaching-learning. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.

Hidi, S. & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*(2), 151–179.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 2*(2), 111–127.

Huff, K. L. & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, *20*(3), 16-25.

Kaufman, G. (2003). What to measure? A new look at the concept of creativity. *Scandinavian Journal of Educational Research. 47*(3), 235-251.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. In L. Idol & B.F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform (pp. 11-40)*. Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking. *Creativity Research Journal, 18*(1), 3-14.

Kouros, C. & Abrami, P. C. (2006). *How do students really feel about working in small groups? The role of student behaviors and attitudes in cooperative classroom settings*. Montreal, Canada: Centre for the Study of Learning and Performance.

Kramarski, B. & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281–310.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335-353.

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(2009), 70–76.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.

Kuhn, D. & Dean, D. (2004). A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*(4), 268–273.

Kuncel, N. R. (2011). Measurement and meaning of critical thinking. Draft report for the National Research Council's 21st century skills workshop, January 2011, Irvine, CA.

Kyllonen, P. (2008). The tried, the true, and the transpiring in creativity assessment. In E. Villalba (organizer), Can creativity be measured? OECD: Brussels, Belgium.

Lange, G. W., & Adler, F. (1997, April). Motivation and achievement in elementary children. Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, D.C.

Lipman, M. (1988). Critical thinking—What can it be? *Educational Leadership, 46*(1), 38–43.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313–327.

Lubart, T. I. & Georgsdottir, A. (2004). Creativity: Developmental and cross-cultural issues. In S. Lau, A.N., Hui, & G.Y. Ng (Eds.), *Creativity: When East meets West* (pp. 23-54). Singapore: World Scientific.

Lubart, T. & Guignard, J. H. (2004). The generality-specificity of creativity: A multivariate approach. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 43-56). Washington, D.C.: American Psychological Association.

Maker, C. J. (2004). Creativity and multiple intelligences: The DISCOVER project and research. In S. Lau, A.N., Hui, & G.Y. Ng (Eds.), *Creativity: When East meets West* (pp. 362-413). Singapore: World Scientific.

Martinez, M. E. (2006). What is metacognition? *Phi Delta Kappan, 87*(9), 696–699.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730-740.

McLeod, L. (1997). Young children and metacognition: Do we know what they know they know? And if so, what do we do about it? *Australian Journal of Early Childhood, 22*(2), 6–11.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(4), 10–12.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359–377.

Miller, S. D., & Meece, J. L. (1997). Enhancing elementary students' motivation to read and write: A classroom intervention study. *The Journal of Educational Research, 90*(5), 286–299.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education, 15*, 363–389.

Moss, P. A., & Koziol, S. M. (1991). Investigating the validity of a locally developed critical thinking test. *Educational Measurement: Issues and Practice*, *10*(3), 17–22.

National Governors Association, Center for Best Practices, and the Council of Chief State School Officers. (2010). Introduction to the Common Core State Standards. Retrieved from http://www.corestandards.org/assets/ccssi-introduction.pdf.

Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher, 18*(9), 21–-26.

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real world problems. *Journal of Psychoeducational Assessment, 9*, 145–153.

Partnership for 21st Century Skills. (2009). P21 framework definitions. Retrieved from http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf.

Partnership for Assessment of Readiness for College and Careers. (2010). Application for the race to the top comprehensive assessment systems competition. Retrieved from http://www.fldoe.org/parcc/pdf/apprtcasc.pdf.

Paris, S. G. & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89-101.

Paris, S. G. & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. *Remedial and Special Education, 11*(6), 7–15.

Paul, R. W. (1992). Critical thinking: What, why, and how? *New Directions for Community Colleges, 1992*(77), 3–24.

Paul, R. W., & Elder, L. (2006). Critical thinking: The nature of critical and creative thought. *Journal of Developmental Education, 30*(2), 34–35.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83-96.

Prins, F. J., Veenman, M. V. J., & Elshout, J. J. The impact of intellectual ability and metacognition on learning: New support for the threshold of problematicity theory. *Learning & Instruction, 16*, 374–387.

Rojas-Drummond, S. & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research, 39,* 99–111.

Roschelle, J. & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem-solving. In C.E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). Berlin: Springer-Verlag.

Runco, M. A. (1993). Divergent thinking, creativity, and giftedness. *Gifted Child Quarterly, 31*(1), 16-22.

Runco, M. A. (2003). Education for creative potential. *Scandinavian Journal of Educational Research, 47*(3), 317-324.

Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55*, 657-687.

Russ, S. W. (1996). Development of creative processes in children. *New Directions for Child Development, 72*, 31-42.

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences, 2*(1), 1–17.

Ryan, R.M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology 24*(1), 54-67.

Salomon, G. & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*(1), 89–100.

Saner, H., McCaffrey, D., Stecher, B., Klein, S., & Bell, R. (1994). The effects of working in pairs in science performance assessments. *Educational Assessment, 2*(4), 325–338.

Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing "intermediate constraint" questions and tasks for technology platforms. *Journal of Technology, Learning, and Assessment, 4*(6). http://escholarship.bc.edu/jtla/vol4/6/

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114-121.

Schneider, W. & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In Perfect, T. & Schwartz, B. (Eds.), *Applied metacognition*. Cambridge, UK: Cambridge University Press.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1-2), 113-125.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139.

Schraw, G. & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371.

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*, 848–856.

SMARTER Balanced Assessment Consortium. (2010). Race to the top assessment program application for new grants. Retrieved from http://www.k12.wa.us/smarter/pubdocs/SBAC_Narrative.pdf.

Smith, G. F. (2002). Thinking skills: The question of generality. *Journal of Curriculum Studies, 34*(6), 659–678.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*, 51-79.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement* National Institute of Education. Retrieved from http://eric.ed.gov/PDFS/ED272882.pdf.

Sternberg, R. J. (2006a). The nature of creativity. *Creativity Research Journal, 18*(1), 87-98.

Sternberg, R. J. (2006b). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence, 34*, 321-350.

Sternberg, R. J. (2010). Teaching for creativity. In R. A. Beghetto & J. C. Kaufman (Eds.), *Nurturing Creativity in the Classroom* (pp. 394–414), New York, NY: Cambridge University Press.

Stipek, D. J. (1996). Motivation and instruction. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 85–113). New York, NY: Macmillan.

Streeter, L., Bernstein, J., Foltz, P., & DeLand, D. (2011). Pearson's automated scoring of writing, speaking, and mathematics. Pearson White Paper. Iowa City, IA: Pearson. Retrieved from http://www.pearsonassessments.com/hai/images/tmrs/PearsonsAutomatedScoringofWritingSpeakingandMathematics.pdf.

Tindal, G. & Nolet, V. (1995). Curriculum-based measurement in middle and high schools: Critical thinking skills in content areas. *Focus on Exceptional Children, 27*(7), 1–22.

Toplak, M. E. & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 94*(1), 197-209.

Torrance, E. P. (1977). *Creativity in the classroom: What research says to the teacher*. Washington, D.C.: National Education Association.

Treffinger, D. J., & Isaksen, S. G. (2005). Creative problem solving: The history, development, and implications for gifted education and talent development. *Gifted Child Quarterly, 49*(4), 342-353.

Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Sarasota, FL: National Research Center on the Gifted and Talented.

Turner, J. C. (1995). The influence of classroom contexts on young children's motivation for literacy. *Reading Research Quarterly, 30*(3), 410–441.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311–330.

Van Tassel-Baska, J., & MacFarlane, B. (2009). Enhancing creativity in the curriculum. In Shavinina, L.V. (Ed.), *International handbook on giftedness* (pp. 1061-1083). Netherlands: Springer.

Watson, G. B., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal manual: Forms A and B.* San Antonio: The Psychological Corporation.

Webb, N.M. (1991). Task-related verbal interaction and mathematical learning in small groups. *Research in Mathematics Education, 22*(5), 366–389.

Webb, N. M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment, 1*(2), 131–152.

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*(2), 239–261.

Webb, N. M., Farivar, S. H., & Mastergeorge, A. M. (2001). *Productive helping in cooperative groups*. CSE Technical Report No. 555. Los Angeles: CRESST.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607–651.

Webb, N. M. & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39,* 73–97.

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review, 12*(1), 1-14.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930-941.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., Almeqdad, Q., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4*(1), 63-85.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator, 31*(2), 8–19.

Zhuang. X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). Development and validity evidence supporting a teamwork and collaboration assessment for high school students. ETS RR-08-50. Princeton, NJ: ETS.