MEASURING STUDENT SUCCESS SKILLS: A REVIEW OF THE LITERATURE ON CRITICAL THINKING

March 31, 2020

21st CENTURY SUCCESS SKILLS

Carla Evans, Ph.D.
National Center for the Improvement of Educational Assessment

Center for Assessment

www.nciea.org

TABLE OF CONTENTS

1. Center for Assessment completed this work on behalf of PBLWorks (Buck Institute for Education) in its effort to provide tools and resources to school and district partners as they assess student success skill performance in Gold Standard Project Based Learning.

BUCK INSTITUTE FOR EDUCATION
PBLWorks

2. I acknowledge the terrific feedback on previous drafts from my colleagues at the Center for Assessment. Any errors and omissions are my own.

4. Evans, C. M. (2020). *Measuring student success skills: A review of the literature on critical thinking.* Dover, NH: National Center for the Improvement of Educational Assessment.

MEASURING STUDENT SUCCESS SKILLS: A REVIEW OF THE LITERATURE ON CRITICAL THINKING

INTRODUCTION

Educational philosophers, from Plato and Socrates to John Dewey, highlighted the importance of critical thinking and the intrinsic value of instruction that reaches beyond simple factual recall (McPeck, 1981). Teaching students to think critically is an avowed aim of education because critical thinking cuts across almost all areas of life (Facione, 1990; Kurfiss, 1988; Paul & Elder, 2007). Critical thinking is vital to "the personal and civic life of all members of society" because it allows individuals to evaluate information presented to them in order to make better judgments (Facione, 1990, p. 32). Critical thinking also provides the foundation for inquiry (Abrami et al., 2008): One cannot investigate claims, evaluate the quality of evidence, or make any other inquiry-based endeavor without thinking critically. Edward Glaser, the father of contemporary research in critical thinking, argued that only citizens with developed critical thinking skills can make intelligent judgments about public issues (Abrami et al., 2015).

Teaching students to think critically is an avowed aim of education because critical thinking cuts across almost all areas of life.

The critical thinking literature is rooted in three fields: psychology, philosophy, and education (Lewis & Smith, 1993; Sternberg, 1986). These disciplines reflect different approaches for defining critical thinking. The *psychological tradition* emphasizes mastery of discrete skills and dispositions that generalize across multiple contexts (Sternberg, 1986); these skills include interpretation, analysis, evaluation, and synthesis. Explicit instruction in critical thinking increases the potential for transfer across contexts (van Gelder, 2005). In this way, the psychological tradition tends to focus on what constitutes good critical thinking.

In contrast, the *philosophical tradition* focuses on the ideal critical thinker, emphasizing the person's qualities and characteristics rather than the behaviors or actions the critical thinker can perform (Lewis & Smith, 1993). Exemplified in the work of Paul and Elder (2007), the philosophical approach traditionally has focused on the use of formal rules of logic applied to content-specific knowledge. This differs from the psychological approach because, philosophers believe, critical thinking skills and dispositions are inseparable from content. Philosophers argue, for example, that no critical thinking skills are necessary or sufficient across all contexts; it depends, rather, on the subject area (Bailin & Siegel, 2003).

Finally, the *educational tradition* of critical thinking stems from the work of Benjamin Bloom. Educators have long relied on Bloom's taxonomy of hierarchical cognitive processing skills for both teaching and assessing higher-order thinking skills. Factual recall and other knowledge-level cognitive processes sit at the bottom of the taxonomy, with the three highest levels—analysis, synthesis, and evaluation—generally seen to constitute critical thinking (Lai, 2011).

Drawing on all three traditions, the purpose of this literature review is to explore the conceptualizations, definitions, and understandings in the research literature related to critical thinking. Key initial questions include: What is critical thinking? How is critical thinking related to other success skill concepts? And to what extent does critical thinking develop over time? This foundational information will then be used to examine (a) instructional approaches to promote critical thinking, (b) benefits of critical thinking on valued student outcomes such as student learning, and (c) ways teachers can collect evidence that reveals the benefits of student critical thinking outcomes using student artifacts and other appropriate measures.

DEFINITIONS

What is Critical Thinking?

There is considerable dispute about how to define critical thinking. This debate relates to the three traditions noted above and their differing conceptualizations of critical thinking. The dispute also stems from different understandings of what it means to think critically. Although all scholars agree that critical thinking pertains to mental processes, they disagree on the extent to which critical thinking includes both cognitive skills and dispositions (Facione, 1990). Researchers note that the *skill* involved in thinking critically is distinct from the disposition to do so (Ennis, 1993). For example, an individual may have the skill to judge the veracity of received information, but not actually employ this skill. In the literature, the most commonly cited critical thinking dispositions are open-mindedness, fair-mindedness, the propensity to seek reason, inquisitiveness, the desire to be well-informed, flexibility, and respect for (and willingness to entertain) others' viewpoints (Lai, 2011).

The American Philosophical Association, in 1990, convened a panel of 46 experts to correspond about how critical thinking should be defined and conceptualized; this exchange resulted in "The Delphi Report" (Facione, 1990). The Delphi panel defined critical thinking as

> *…purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. . . .The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider . . . and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit (Facione, 1990, p. 3).*

The Delphi panel defined critical thinking as *…purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation…*

As summarized in Table 1, the Delphi panel identified six cognitive skills (and corresponding subskills) and two dispositions (and corresponding sub-dispositions), which, together, provide a framework for understanding and assessing critical thinking. The National Research Council (2012) concurred with The Delphi Report, agreeing that critical thinking entails both cognitive skills and dispositions.

Table 1.
List of Critical Thinking Skills and Dispositions from The Delphi Report

Category	Description
Cognitive skills and subskills	**Interpretation:** • Categorization • Decoding significance • Clarifying meaning
	Analysis: • Examining ideas • Identifying arguments • Analyzing arguments
	Evaluation: • Assessing claims • Assessing arguments
	Inference: • Querying evidence • Conjecturing alternatives • Drawing conclusions
	Explanation: • Stating results • Justifying procedures • Presenting arguments
	Self-Regulation: • Self-examination • Self-correction
Dispositions and sub-dispositions	**Approaches to specific issues, questions, or problems:** • Clarity in stating the question or concern • Orderliness in working with complexity • Diligence in seeking relevant information • Reasonableness in selecting and applying criteria • Care in focusing attention on the concern at hand • Persistence though difficulties are encountered • Precision to the degree permitted by the subject and the circumstance
	Approaches to life and living in general: • Inquisitiveness with regard to a wide range of issues • Concern to become and remain generally well-informed • Open-mindedness regarding alternatives and opinions • Understanding of the opinions of other people • Fair-mindedness in appraising reasoning • Honesty in facing one's own divergent world views • Flexibility in considering biases, prejudices, stereotypes, egocentric or sociocentric tendencies • Prudence in suspending, making or altering judgments • Willingness to reconsider and revise views where honest reflection suggests that change is warranted

Note. See Facione (1990, pp. 6-13) for a more complete description of each skill and disposition.

Are Critical Thinking Skills Generic or Discipline-Specific?

Are critical thinking skills general, transferable, and applicable across disciplinary areas, or, rather, does critical thinking differ according to the content domain and context in which it is taught and applied? This is an important question, to be sure. If critical thinking is generic, then it arguably could be taught independently in separate courses, with the sole focus being on the development of critical thinking skills. But if critical thinking is regarded as particular to a discipline, then it should be taught embedded within subject-matter content.

There are at least three positions on this question. The *generalist position* (Siegel, 1991; Pithers & Soden, 2000) contends that many aspects of critical thinking, such as identifying faulty reasoning and informal fallacies, are generalizable across disciplines. According to Siegel (1991), critical thinking is generalizable in this regard because errors of reasoning are based on argument design, not content. And while Pithers and Soden (2000) concede that critical thinking may have subject-specific elements, "it is difficult to conceive of any broad type of thinking that has no significant application outside a particular discipline" (p. 246). For example, the basic building blocks of inquiry and research, such as forming and testing hypotheses, seem sufficiently broad that they need not be learned anew in each discipline, but, rather, can be transferred across contexts.

The *specifist position* (McPeck, 1981) argues for the indispensability of content knowledge in the critical thinking process. McPeck dismisses the generalist's argument for three reasons: (a) all thinking is thinking about something; (b) general critical thinking ability is not possible because knowledge of a subject is necessary for critical thinking; and (c) critical thinking varies greatly from discipline to discipline. Critical thinking in biology, for example, may well be different from critical thinking in history or art.

The *blended position*, in short, is a mixture of the two (Ennis, 1993; Facione, 1990; Kurfiss, 1988; Paul & Elder, 2007). While recognizing that critical thinking skills can be distinguished from content knowledge, adherents of the blended position assert that "one of the best ways to learn critical thinking is within a subject area" (Facione, 1990, p. 32). Additionally, because what counts as evidence vary from discipline to discipline, general critical thinking skills are necessary but insufficient for enabling critical thought within a given discipline (Kurfiss, 1988; Lai, 2011).

What is the Relationship between Critical Thinking and other Success Skill Concepts?

Not every important thinking skill is critical thinking. The Delphi Report authors contend that "critical thinking is one among a family of closely related forms of higher-order thinking, along with, for example, problem solving, decision making, and creative thinking" (Facione, 1990, p. 16). Although these concepts are related and overlap in practice, they nevertheless have distinguishing characteristics.

Not every important thinking skill is critical thinking.

Problem solving, for example, involves focusing on a problem and finding a solution—a narrow focus and scope. Critical thinking goes beyond problem solving because critical thinking is not necessarily centered on finding a solution (Simpson & Courtney, 2003). Critical thinking is broader, focusing on interrogating assumptions, evaluating claims, and critiquing solutions based on evidence.

Many researchers have connected creativity and critical thinking (Lai, 2011). Paul and Elder (2006), for example, see creativity and critical thinking as two sides of the same coin. Good critical thinking requires the ability to create intellectual products, which relates to creativity and creative thinking. Good critical thinking also requires the individual to analyze and evaluate the quality, usefulness, and sufficiency of those intellectual products, which relates to specific cognitive skills and dispositions. As Paul and Elder argue,

"Critical thinking without creativity reduces to mere skepticism and negativity, and creativity without critical thought reduces to mere novelty" (p. 35). In practice, the two are inextricably linked and develop in parallel.

In the National Research Council (2012) report, the authors consider critical thinking, along with creative thinking and content knowledge, to fall in the cognitive competency category. Yet critical thinking is also related to the other two categories—intrapersonal and interpersonal competencies. The authors of the report presented a range of intrapersonal and interpersonal competencies that are closely related and often employed in tandem as students or individuals transfer their learning to new or novel situations. Table 2 summarizes these competencies.

Table 2.
National Research Council 21st Century Competencies

Cognitive Competencies	Intrapersonal Competencies	Interpersonal Competencies
Cognitive Processes and Strategies Critical thinking, problem solving, analysis, reasoning/ argumentation, interpretation, decision making, adaptive learning, executive function	**Intellectual Openness** Flexibility, adaptability, artistic and cultural appreciation, personal and social responsibility	**Teamwork and Collaboration** Communication, collaboration, etc.
Knowledge Information literacy, information and communications technology literacy, oral and written communication, active listening	**Work Ethic/ Conscientiousness** Motivation and self-direction, responsibility, Type 1 self-regulation (metacognition, including forethought, performance, and self-reflection), and perseverance	**Leadership**
Creativity Creativity and innovation	**Positive Core Self Evaluation** Type 2 self-regulation (self-monitoring, self-evaluation, self-reinforcement), physical and psychological health	

Note. Figure adapted from National Research Council, 2012.

Metacognition—"thinking about thinking"—also is involved in the process of critical thinking. Metacognition can be seen as a supporting condition for critical thinking, in that monitoring the quality of one's thought makes it more likely that one will engage in good critical thinking (Lai, 2011). Lai (2011) argues further that self-regulation provides the link between the two. Self-regulation is the ability to plan, direct, and control one's emotions, thoughts, and behaviors during a learning task. In short, students use self-regulation as they monitor the quality of their thought (metacognition), which then supports critical thinking.

Critical thinking is related to motivation as well. As Facione (2000) put it, the disposition to think critically reflects the "consistent internal motivation to engage problems and make decisions" (p. 65). Motivation, like metacognition, is viewed as a supporting condition for critical thinking in that unmotivated individuals are unlikely to exhibit good critical thinking (Lai, 2011).

DEVELOPMENT

How Does Critical Thinking Develop?

Adults do not always employ critical thinking when it's called for. Climate change, for example, is a controversial topic because people are not necessarily swayed by facts, reason, or evidence; many find personal experience more compelling than logical thought or empirical evidence. Given this natural tendency, Halpern (1998) warns that we should not expect dramatic improvements in critical thinking as a result of instructional interventions; improvement, rather, likely will be slow and incremental.

Research suggests there is no single age when children are developmentally ready to learn more complex ways of thinking.

People begin developing critical thinking competencies at a very young age, and the efficacy of instructional interventions for improving critical thinking does not differ by grade level (Abrami et al., 2015). This finding is surprising from a Piagetian perspective, which views the cognitive processes of young children as undeveloped compared with those of older individuals. However, research suggests there is no single age when children are developmentally ready to learn more complex ways of thinking (Silva, 2008), a finding consistent with both sociocultural and cognitive learning theory.

The Delphi Report recommends that, "from early childhood, people should be taught . . . to reason, to seek relevant facts, to consider options, and to understand the views of others" (Facione, 1990, p. 27). However, little is known about how critical thinking skills and dispositions develop (Lai, 2011); there are no learning progressions of critical thinking skills and dispositions. Indeed, the Delphi Report cautioned that its framework for critical thinking should not be interpreted as implying a developmental progression or hierarchical taxonomy (Facione, 1990). Only one researcher has published a developmental progression of critical thinking skills and dispositions (Kuhn, 1999)—a perspective arguably at odds with sociocultural and cognitive perspectives, which do not restrict when students are able to start learning to think deeply.

INSTRUCTION

What Are Some Instructional Approaches to Teaching Critical Thinking?

The debate regarding general versus content-specific critical thinking skills, discussed above, has considerable implications for how teachers approach the teaching of critical thinking.

The debate regarding general versus content-specific critical thinking skills, discussed above, has considerable implications for how teachers approach the teaching of critical thinking.

If critical thinking is generic, as argued earlier, then it can be taught independently in separate courses, focusing solely on critical thinking development. If critical thinking is content and context dependent, however, then its instruction should be embedded in the discipline.

Ennis (1989) provided a typology of critical thinking approaches for classifying and describing various instructional interventions related to teaching critical thinking: general, immersion, and mixed approaches. These typologies are aligned with the generalist, specifist, and blended perspectives respectively.

Table 3 summarizes the three instructional approaches related to teaching critical thinking and identifies the extent to which general critical thinking is explicitly or implicitly taught, and the extent to which subject matter instruction is woven into each approach. In the *general* approach, generic critical thinking skills and dispositions are explicitly taught and are the learning objectives, without specific subject matter content. In contrast, students are immersed in subject matter content knowledge without instruction in generic critical thinking in the *immersion* approach. The *mixed* approach typically involves a separate section of the course that explicitly teaches general critical thinking skills and the rest of the course embeds critical thinking instruction within subject-specific norms and evidentiary arguments.

Table 3.
Types of Talk Activated in Collaborative Learning Activities from Less to More Sophisticated

Approach *(Position)*	General Critical Thinking Skills (Explicitly or Implicitly Taught)	Subject Matter Instruction (High or Low)
General ***(Generalist)***	Explicit	Low
Immersion ***(Specifist)***	Implicit	High
Mixed ***(Blended)***	Explicit	High

What Do We Know About the Effects of Instruction on the Development of Critical Thinking Skills and Student Achievement?

Abrami et al. (2008, 2015) conducted two meta-analyses to synthesize the empirical research on the effects of instruction on the development and enhancement of critical thinking skills and dispositions, the latter of which also included effects on student achievement. Overall, findings from hundreds of experimental or quasi-experimental studies show that instruction improves critical thinking skills and dispositions, corresponding to an average effect size of +.30, or almost one third of a standard deviation. This effect size is small to moderate by conventional standards for judging this statistic, but practically meaningful in the educational research literature and similar to the impact of formative assessment on student achievement (Black & Wiliam, 1998).

Overall, findings from hundreds of experimental or quasi-experimental studies show that instruction improves critical thinking skills and dispositions, corresponding to an average effect size of +.30, or almost one third of a standard deviation.

This instructional effect, however, is neither uniform nor consistent across studies. For example, the authors' 2008 meta-analysis revealed larger effects for K-12 students than for undergraduates (although this finding was not replicated in their later meta-analysis). Further, the 2008 study found that the type of critical

thinking intervention also explained variability in instructional effects. Mixed-approach interventions—those combining both subject-matter content and explicit critical thinking instruction—significantly outperformed all other types of instruction, whereas immersion interventions significantly underperformed all other approaches. That is, the least effective approach was to immerse students in thought-provoking subject matter instruction without explicit use of critical thinking principles.

In their 2015 meta-analysis, Abrami et al. found that two general types of instructional interventions are key to the development of critical thinking skills: (a) using collaborative or cooperative learning methods; and (b) using real-life problems, situations, and examples. Many of the studies included in this later meta-analysis contained measures of course-content learning (e.g., chemistry) as well as generic critical thinking skills. Most of these measures were teacher-made, so their psychometric properties remain unknown. That said, the average effect size was +.33, with, as in the authors' earlier meta-analysis, individual effects varying considerably across studies.

In terms of teacher training, findings across studies suggest that professional development in teaching critical thinking skills had more of an impact on teacher efficacy than teachers observing other teachers, receiving detailed curriculum guides, or listing critical thinking objectives as part of course objectives (Abrami et al., 2008).

These findings show that improving students' critical thinking skills and dispositions cannot be left simply to implicit expectations. Rather, teachers must make critical thinking objectives explicit in their lessons and courses, and critical thinking instruction must be addressed in both preservice and in-service training. Findings also suggest there are effective strategies for improving critical thinking skills, such as dialogue/question posing and authentically situated instruction. Teachers can facilitate opportunities for students to engage in critical thinking activities as they provide real-world opportunities for solving problems with multiple solutions, provide structure that allows students to respond to open-ended questions and formulate and articulate solutions to problems, and provide a variety of learning activities that allow students to choose and engage in solving authentic problems.

MEASUREMENT/ASSESSMENT

How is Critical Thinking Typically Measured or Assessed?

Measuring critical thinking is complex. Critical thinking has elements that are both domain-specific and generic. Critical thinking also involves cognitive skills *and* dispositions. Researchers and practitioners have used two main types of instruments for assessing critical thinking: standardized tests and performance-based assessments.

Standardized Tests

There are many standardized tests of critical thinking skills, such as the Watson-Glaser Critical Thinking Appraisal test and California Critical Thinking Skills Test. There are also a few for assessing critical thinking dispositions, such as the California Critical Thinking Dispositions Inventory. And then there are the recent computer-based assessments of critical thinking, such as Educate Insight's Reasoning Skills and Thinking Mindset for Grades K-2, 3-5, 6-8, and 9-12 (The California Academic Press, n.d.). The Reasoning Skills assessment targets the cognitive domain of critical thinking, and the Thinking Mindset the affective domain (both using multiple-choice questions).

Construct underrepresentation is the inherent weakness of many standardized tests. This is particularly relevant to critical thinking, as many of the existing measures focus on generic critical thinking skills. Additionally, a multiple-choice format may not adequately capture the dispositional characteristics of

test-takers. For example, Ku (2009) argues that multiple-choice critical thinking tests only measure recognition and, therefore, cannot "reveal test-takers' underlying reasoning for choosing a particular answer . . . [or] reflect test-takers' ability to think critically under unprompted situations" (p. 70).

Performance-Based Assessments

High-quality performance-based assessments require students to apply (or transfer) their knowledge and skills to novel contexts. Performance assessments involve students producing something (e.g., report, product, experiment or demonstration), which is then evaluated against specific criteria found in a rubric or scoring guide. Such measures are well-suited to gather evidence of students' level of sophistication in applying critical thinking skills and dispositions.

Performance assessments are well-suited to gather evidence of students' level of sophistication in applying critical thinking skills and dispositions.

Several organizations and researchers have created critical-thinking rubrics to accompany performance-based assessment or project-based learning (American Association of Colleges and Universities, 2009; Buck Institute for Education, 2019; Saxton, Belanger, & Becker, 2012; Washington State Career and Technical Education, n.d.). Reflecting the Delphi Report's definition of critical thinking, the Critical Thinking Analytic Rubric (CTAR; Saxton et al., 2012) comprises six dimensions—interpretation, analysis, evaluation, inference, explanation, and disposition—and is scored on a 6-point scale. Acceptable levels of both intra- and inter-rater reliability have been reported, although no evidence regarding validity can be found.

What are the Measurement/Assessment Issues Related to Critical Thinking?

There are instructional, practical, and technical considerations when selecting (or designing) measures of 21st century competencies (Soland et al., 2013). Instructional considerations pertain to the use of assessment information. For example, is the measure intended to be used formatively or summatively? Is it to provide actionable information to teachers, or useful feedback to students? Is the assessment grade, context, or culturally appropriate? Practical considerations relate to cost and ease of administration, delivery, and scoring. And technical considerations center on validity, reliability, and fairness.

The desired inferences that educators wish to make from assessment results will influence what evidence will be collected (NRC, 2011; Wilson et al., 2012). The development of educational and psychological tests typically proceed as follows: define the targeted construct; create tasks to elicit desired responses; select item types; consider the various administration issues; determine the values, codes, or scores to be assigned to student responses; pilot the assessment, using a large and diverse sample of students; model and analyze responses, attending to technical issues such as validity, reliability, and test fairness.

The desired inferences that educators wish to make from assessment results will influence what evidence will be collected.

With respect to defining the construct of critical thinking, the desire to measure both cognitive skills and dispositions requires close attention to test format (Ku, 2009). Multiple-choice items may adequately measure generic critical thinking skills, but may not adequately capture the dispositional aspects of critical thinking. Students can select responses to items that elicit analysis, for example, but that selection does not mean that a student is likely to apply analysis skills in context.

Additionally, because the debate over domain specificity in critical thinking remains unresolved, assessing critical thinking is difficult because the type of inferences (or claims) that can be made from assessment results is unclear (Lai, 2011). For example, if students demonstrate critical thinking skills in a particular subject area, would they be able to transfer those skills to another topic in the same subject area (near transfer) or to another subject area altogether (far transfer)? When students do not transfer critical thinking skills to another subject area, is this because they need additional instruction in critical thinking, additional instruction in the subject area, or both? Not only are the cognitive aspects of critical thinking confounded with subject-specific knowledge, but the dispositional aspects of critical thinking are confounded with the ability to think critically (Lai, 2011). Just because researchers agree that critical thinking comprises both cognitive skills and dispositions does not mean that it is possible to delineate their separate effects in practice (Lai, 2011).

What are the Implications of Research for Assessment Design and Use?

Assessment Design

The following implications for assessment design are related to classroom-based assessments in particular. These implications are not limited to critical thinking, as they also apply to most of the other student success skills.

While some students may exhibit critical thinking without being prompted, most will rise or sink to what the task requires.

First, ***assessment tasks should prompt complex judgments.*** While some students may exhibit critical thinking without being prompted, most will rise or sink to what the task requires. The materials (visual, texts, etc.) used to elicit students' critical thinking therefore are crucial and have a sizable impact on the extent to which critical thinking is elicited in any given assessment experience. In other words, if the task doesn't ask students to think critically, they likely will not demonstrate evidence of critical thinking. The task, embedded in projects or other curriculum activities, must be designed and structured thoughtfully to elicit students' critical thinking.

Similarly, ***assessment tasks should include open-ended and/or ill-structured tasks.*** Open-ended tasks are the opposite of standardized assessments, which rely heavily on selected response item types that assess limited aspects of critical thinking and other 21st century skills (Ku, 2009; Lai & Viering, 2012). Open-ended tasks allow students to decide what information is relevant, how to use the information, and how to demonstrate their understanding of the information; open-ended tasks also allow multiple solution pathways. In contrast, closed tasks typically have one correct solution, and the teacher indicates what information is relevant and how the information is to be presented. An ill-structured task has "no clearly defined parameters, no clear solution strategies, and either more than one correct solution, or multiple ways, of arriving at an acceptable solution" (Lai & Viering, 2012, p. 46). Fischer, Spiker, and Riedel (2009) found that stimulus material that is conflicting, disordered, and/or uncertain generated more critical thinking in U.S. Army officers than did consistent and coherent stimulus materials (p. vi). The advantages of open-ended and/or ill-structured tasks for the purpose of measuring critical thinking is that such tasks allow for multiple, defensible solutions and, further, require students to apply (or transfer) their learning to novel situations.

Assessment tasks should be authentic. As Care et al (2018) state: "the premise for good assessment is that it captures valid indicators of the target construct…to stimulate the behaviors from which these indicators can be captured, the assessment design must mirror the real-life demands of a situation that would provoke

those behaviors" (p. 20). Authentic, real-life contexts do not guarantee the validity of the assessment information for any particular use, but authenticity does contribute to validity. Similarly, authenticity is an important consideration for student motivation and engagement—both of which also relate to validity.

Finally, ***assessment tasks should make student thinking visible to teachers.*** To provide formative feedback regarding the quality of students' critical thinking, teachers must have assessment tasks that render student thinking visible. This can be accomplished in multiple ways, but their commonality is that all approaches likely will require students to provide written or verbal evidence that support their claims, judgments, assertions, and so on.

Assessment Use

There are many challenges with assessment use regarding 21st century skills. First and foremost, there are no clear end of grade-level or grade-span standards that define proficiency for any of the success skills, including critical thinking. There is at least one research-based, hypothesized learning progression of how students demonstrate less to more sophisticated forms of critical thinking in the literature (Quinn, McEachen, Fullan, Gardner, & Drummy, 2020). This learning progression is analytic and multi-dimensional, with five levels of student performance, and describes performance in grades K-12 (i.e., broken down by neither grade level nor grade span).

It is unclear how students develop competence in the domain of critical thinking, and there are no expected levels of critical thinking at certain markers in time.

Empirically validated learning progressions do not yet exist for student success skills. Consequently, it is unclear how students develop competence in the domain of critical thinking, and there are no expected levels of critical thinking at certain markers in time. It also is unclear what exactly (if anything) becomes more complex over time related to critical thinking skills. Is it the case that critical thinking skills, such as analysis, become more sophisticated over time? Or is it that the assessment tasks and disciplinary content to which students are applying these skills become more complex (or novel) over time? Or is it a combination of both?

An additional challenge with assessment use relates to the creation of rubrics to score and grade student performance in any particular student success skill. Rubrics imply scoring and grading, and grading can have negative effects on student learning (Shepard, 2019). This is because grading can elicit comparisons among students, which can adversely affect student motivation. More, specifically, grading 21st century skills is fraught with potential unintended consequences, as the measures are not sufficiently accurate at the individual student level and distort the meaning of grades as indicators of academic achievement. Additionally, there is a long and deep research base related to assessment for learning and how students learn more from written formative feedback than from grades (Black & Wiliam, 1998).

For these reasons, we suggest not using the language of a rubric, but instead creating research-based continua to describe student performance from less to more sophisticated. The Delphi Report (Facione, 1990) provides one normative framework for creating critical thinking draft continua. The six cognitive skills and two groupings of dispositions that constitute the Delphi definition of critical thinking could be used for structuring either subject-specific analytic continua or generic critical thinking continua having subject-specific annotations. This approach aims to balance the domain-specific and generic nature of critical thinking. These continua would be pilot tested on student work in local contexts to evaluate the extent to which they accurately reflect how students across socio-cultural contexts and conditions demonstrate competence in the domain.

Additionally, the purpose of the continua would be to provide useful, formative information that teachers could use to guide instruction and provide feedback to students on the quality of their critical thinking. The student work analysis could provide a way to evaluate the extent to which the continua provide valuable and useful feedback to students, parents, and teachers for instructional purposes. Annotated student work samples from across disciplines and types of assessment tasks would be especially useful in helping teachers recognize markers for the essential elements of critical thinking in student work products and artifacts.

Because the so-called deeper learning competencies (cognitive, interpersonal, and intrapersonal) are intertwined, assessments intended to elicit evidence of a student critical thinking skill will also elicit evidence of student's ability to, say, communicate (verbal/written expression) and keep themselves on task (self-directed learning). The interrelations among student success skills may necessitate a more holistic and complex understanding of student competence within and across content areas. This has implications for the use of assessment information for any student success skill, including critical thinking.

Finally, students should be given multiple opportunities to demonstrate critical thinking in various subjects throughout the year. The research evidence suggests that while critical thinking is instructionally sensitive, it does not necessarily change over the short term. Consequently, it would be prudent to collect a body of evidence over the course of the year (or even better over the course of a student's high school experience) to support any kind of generalizable claim about students' critical thinking. The literature remains unclear, however, regarding how much growth in students' critical thinking we should expect to see from one year to the next and, further, if there are particular patterns or stages of critical thinking development that a continua could incorporate.

Students should be given multiple opportunities to demonstrate critical thinking in various subjects throughout the year.

CONCLUSION

The purpose of this paper was to conceptualize and describe critical thinking from the educational, philosophical, and psychological literature; synthesize research findings; and discuss implications for assessment design and use. Overall, findings suggest critical thinking involves both cognitive skills and dispositions. These two aspects are captured in a consensus definition reached by a panel of leading critical thinking scholars and researchers. The resulting Delphi Report defines critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1999, p. 3).

Debate continues about the extent to which critical thinking is generic or domain-specific and, further, how critical thinking develops. That said, research suggests that even young children demonstrate aspects of critical thinking. Empirical research also shows that critical thinking can be taught and that there are specific instructional approaches and strategies that promote more critical thinking. These instructional approaches include explicit teaching of subject matter content within a course that also teaches critical thinking skills. Instructional strategies that promote critical thinking include providing (a) opportunities for students to solve problems with multiple solutions, (b) structure that allows students to respond to open-ended questions and formulate solutions to problems, and (c) a variety of learning activities that allow students to choose and engage in solving authentic problems.

Assessing 21st century skills such as critical thinking is challenging. Educators must attend to how the assessment design prompts, or elicits, student's critical thinking. Assessments must be thoughtfully designed and structured to prompt complex judgments; include open-ended, ill-structured tasks that allow for multiple, defensible solutions; engage students in authentic, real-world scenarios; and make student reasoning visible to teachers. At potential odds with instructional goals is creating critical thinking rubrics to score and grade students. Given the lack of empirical evidence related to how students should develop competence in the domain of critical thinking by the end of some period of time (end of grade, grade span, or 12th grade), we recommend that draft critical thinking continua be created to describe student performance from less to more sophisticated, using shared markers of critical thinking skills. We recommend these draft continua be tested and evaluated against student work to ascertain the accuracy of their descriptions of student performance and usefulness for teaching and learning purposes in K-12 classrooms.

REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. https://doi.org/10.3102/0034654308326084

American Association of Colleges and Universities. (2009). Critical thinking VALUE rubric. Retrieved from https://www.aacu.org/value/rubrics/critical-thinking

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The Blackwell Guide to the Philosophy of Education* (pp. 181–193). Blackwell Publishing.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–73.

Buck Institute for Education. (2019). Critical thinking rubric for PBL. Retrieved from https://my.pblworks.org/resource/document/3_5_critical_thinking_rubric_ccss_aligned

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher,* 18, 4–10.

Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice, 32*(3), 179–186. Retrieved from https://www.jstor.org/stable/1476699

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Newark, DE: American Philosophical Association.

Fischer, S. C., Spiker, V. A., & Riedel, S. L. (2009). Critical thinking training for army officers volumen two: A model of critical thinking (Research Report 1882). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(1), 70–76. https://doi.org/10.1016/j.tsc.2009.02.001

Kuhn, D. (1999). A developmental model of critical thinking. Educational Researcher, 28(2), 16–25, 46. Retrieved from https://www.jstor.org/stable/1177186

Kurfiss, J. G. (1988). Critical thinking: Theory, research, practice, and possibilities. Washington, DC: Association for the Study of Higher Education. Retrieved from http://eric.ed.gov/?id=ED304041

Lai, E. (2011). Critical thinking: A literature review. Pearson. http://www.pearsonassessments.com/research

Lai, E., & Viering, M. (2012). Assessing 21st century skills: Integrating research findings. In *Annual Meeting of the National Council for Measurement in Education.* Vancouver, B.C.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice, 32*(3), 131–137.

McPeck, J. E. (1981). *Critical thinking and education.* London, UK: Routledge.

National Research Council. (2012). *Education for life and work: developing transferable knowledge and skills in the 21st century.* (J. W. Pellegrino & M. L. Hilton, Eds.). Washington, DC: National Academies Press.

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

Paul, R., & Elder, L. (2007). A guide for educators to critical thinking competency standards: Standards, principles, performance indicators, and outcomes. Foundation for Critical Thinking Press. Retrieved from http://www.criticalthinking.org/files/SAM_Comp Stand_07opt.pdf

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: a review. *Educational Research, 42*(3), 237–249. https://doi.org/10.1080/001318800440579

Quinn, J., McEachen, J., Fullan, M., Gardner, M., & Drummy, M. (2020). *Dive into deep learning: Tools for engagement.* Thousand Oaks, CA: Corwin.

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing, 17*(4), 251–270.

Silva, E. (2008). Measuring skills for the 21st century. Washington, DC: Education Sector.

Simpson, E., & Courtney, M. (2003). Critical thinking in nursing education: Literature review. *International Journal of Nursing Practice, 8*(2), 89–98. https://doi.org/10.1046/j.1440-172x.2002.00340.x

Sternberg, R. J. (1986). Critical thinking: Its nature, measurement, and improvement. Washington, DC: National Institute of Education.

The California Academic Press. (n.d.). Educate INSIGHT: K-12 School Assessments. Retrieved from https://www.insightassessment.com/Products/Products-Summary/EDUCATE-INSIGHT-K-12-School-Assessments

van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 53*(1), 41–46. https://doi.org/10.3200/CTCH.53.1.41-48

Washington State Career and Technical Education. (n.d.). 21st Century Skills Rubric: Critical Thinking and Problem Solving.

Wilson, M., Bejar, I., Scalise, K., Templin, J., Wiliam, D., & Irribarra, D. T. (2012). Perspectives on methodological issues. In P. Griffin (Ed.), *Assessment and Teaching of 21st Century Skills.* Springer. https://doi.org/10.1007/978-94-007-2324-5

MEASURING STUDENT SUCCESS SKILLS: A REVIEW OF THE LITERATURE ON CRITICAL THINKING

Center for Assessment

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

www.nciea.org