Pearson

Skills for Today:

What We Know about Teaching and Assessing Critical Thinking

Illustration by Lucy Vigrass

P21 PARTNERSHIP FOR 21ST CENTURY LEARNING

Written by

Matthew Ventura

Emily Lai

Kristen DiCerbo

About the Authors

Dr. Matthew Ventura is the Director of Knowledge and Skills in Education Research at Pearson. Matthew leads research around defining and assessing knowledge and skills in Pearson products. Prior to joining Pearson, Matthew was a senior research scientist at Florida State University where he developed simulations and educational games to assess and support twenty-first-century skills. Matthew received his Ph.D. in Cognitive Science in 2005 from the University of Memphis.

Dr. Emily Lai is Director of Formative Assessment and Feedback in Education Research at Pearson. In that capacity, she leads a research agenda around assessment for learning and principles of effective feedback, particularly within digital environments. Her interests include principled assessment design approaches, such as evidence-centered design, performance assessment, and assessment of twenty-first-century competencies. Emily holds a Ph.D. in educational measurement and statistics from the University of Iowa, a master's degree in library and information science from the University of Iowa, and a master's degree in political science from Emory University.

Dr. Kristen DiCerbo is Vice President of Education Research at Pearson. She leads a team of researchers focused on conducting and translating research about learners and learning in order to influence the development of curricula and digital tools. Her personal research program centers on interactive technologies, particularly the use of evidence from learner activity in games and simulations, to understand what learners know and can do. Prior to joining Pearson, Kristen provided research support to the networking academies at Cisco and was a school psychologist in a local school district in Arizona. Kristen received her master's degree and Ph.D. in educational psychology at Arizona State University.

About Pearson

Pearson is the world's leading learning company. Our education business combines 150 years of experience in publishing with the latest learning technology and online support. We serve learners of all ages around the globe, employing 45,000 people in more than seventy countries, helping people to learn whatever, whenever and however they choose. Whether it's designing qualifications in the UK, supporting colleges in the United States, training school leaders in the Middle East or helping students in China learn English, we aim to help people make progress in their lives through learning.

About P21

P21 recognizes that all learners need educational experiences in school and beyond, from cradle to career, to build knowledge and skills for success in a globally and digitally interconnected world. Representing over 5 million members of the global workforce, P21 is a catalyst organization uniting business, government and education leaders from the United States and abroad to advance evidence-based education policy and practice and to make innovative teaching and learning a reality for all.

Introduction to the Series

This paper is the second in a series to be jointly released by Pearson and P21 entitled, "Skills for Today." Each paper summarizes what is currently known about teaching and assessing one of the Four Cs: collaboration, critical thinking, creativity, and communication. Our partnership on this series signifies a commitment to helping educators, policy-makers, and employers understand how best to support students in developing the skills needed to succeed in college, career, and life.

Sample reference: Ventura, M., Lai, E., & DiCerbo, K. (2017). Skills for Today: What We Know about Teaching and Assessing Critical Thinking. London: Pearson.

table of contents

Foreword

As both of us reflect on our jobs over the years, it is clear that critical thinking has been a consistent demand in all our positions. The skills Dave needed to write a grant proposal at P21 were nearly identical to those required to design the structure of the National Faculty at the Buck Institute for Education. Critical thinking has been at the center of every work engagement for Leah including crafting a strategy for the employability group at Pearson to evaluating the viability of new products to fund. Skills around analysis and persuasive writing have all been clearly important to our success within major organizations.

To be honest, the education and business communities have been talking about the importance of critical thinking for a number of years now. Educators have been espousing the benefits of critical thinking while the business community has been expressing disappointment in the critical-thinking skills of incoming employees. As pointed out in the paper, this seeming contradiction is likely because there is no agreed-upon consensus on how to support and improve critical thinking in specific disciplines.

Despite differing definitions and points of view, progress has been made in investigating the best ways to teach and assess critical thinking skills. Technology, including the possibility of automated scoring around simulations, has opened doors for assessment. At the same time, basic ideas such as using concept maps have been shown to be successful in improving critical-thinking skills. We are excited that Pearson and P21 can partner to produce this summary of the current state of the field.

Of course, critical thinking isn't the only personal and social capability of importance. This is the second of four papers that P21 and Pearson will release. The first was on collaboration and papers on communication and creativity will follow. In order to see large scale change in student proficiency on these, teachers from K through college will need to find ways to teach these skills in their already full schedules of content. We will need to find ways to support them in this process, and hope both the business and policy communities can contribute to these solutions. Guiding students toward becoming good critical thinkers in both their work and civic lives will depend on both local and system-level change.

Leah Jewell, Managing Director, Career Development and Employability, Pearson, and David Ross, CEO, P21

Image by Santanu Das

Introduction

Critical-thinking skills are essential for positive participation in our society. Individuals must be able to make informed decisions based on incomplete or misleading information. Given this requirement, it is probably not surprising that critical thinking is one of the most frequently discussed skills in education, believed to play a central role in logical thinking, decision-making, argumentation, and problem-solving (Butler et al., 2012; Ennis, 1985; Facione, 1990; Halpern, 2003). Broadly, critical thinking is defined as a multifaceted skill that involves problem-solving in the face of ill-defined information. (See the next section for a detailed definition.)

Critical-thinking skills are also proposed to lead to better outcomes in life. For example, among college students and adults in the community, a standardized measure of critical-thinking skills was significantly related to a set of real-world, interpersonal, business, and financial outcomes (Butler et al., 2012). People with higher critical-thinking skills were less likely to report experiencing negative real-world outcomes, ranging from "returning a movie you rented without having watched it at all" to "paying rent or mortgage payment at least two months too late" or "receiving a DUI for drunk driving." Students who receive critical-thinking instruction are more willing to accept scientifically based theories (Rowe et al., 2015), and greater critical-thinking skills are related to greater political participation (Guyton, 1988).

Further, 95 percent of the chief academic officers from 433 higher-education institutions rated critical thinking as one of the most important skills for students to acquire (Association of American Colleges and Universities, 2011). The interest in critical thinking also extends to international institutions and organizations. For instance, the Organisation for Economic Co-operation and Development (OECD, 2012) includes critical thinking as a core skill in college students across the world. And research has found that performance in critical thinking predicts college GPA (grade point average; ACT, no date).

The importance of critical thinking is echoed in the workforce, where 81 percent of employers surveyed wanted colleges to place a stronger emphasis on critical thinking (Association of American Colleges & Universities, 2011). Similarly, a survey given to 400 employers found 92 percent identified critical thinking as one of the top skills needed in college graduates (Casner-Lotto & Barrington, 2006). In interviews with leaders at 200 companies, critical thinking was one of the most frequently mentioned skills essential for both academic and career success (Educational Testing Service, 2013).

This white paper will first give a brief history of the evolution of definitions around critical thinking. Then it will review research on methods for teaching critical-thinking skills among primary, secondary, and post-secondary audiences. We will end with a discussion of how to effectively assess critical-thinking skills. Along the way, we will address key controversies, such as whether critical thinking is better thought of as a general skill or as a domain-specific skill. This white paper will hopefully provide clarity around how to foster critical thinking by providing guidance around activity creation that can elicit critical thinking.

Definitions and Models

Existing Critical-Thinking Frameworks

ACADEMIC MODELS

There are many schools of thought around critical thinking going back to the 1950s. Benjamin Bloom and his associates (1956) created a taxonomy for information-processing skills. Bloom's taxonomy is hierarchical, with "knowledge" at the bottom and "evaluation" at the top. The three highest levels (analysis, synthesis, and evaluation) are frequently said to represent critical thinking (Kennedy, Fisher, & Ennis, 1991). Bloom's is one of the most widely cited sources for educational practitioners when it comes to teaching and assessing higher-order thinking. The benefit of this framework is that it is based on years of classroom experience and observations of student learning (Sternberg, 1986). However, some have noted that the taxonomy lacks the clarity necessary to guide instruction and assessment in a useful way (Ennis, 1985; Sternberg, 1986).

Facione (1990) led a large effort to define critical thinking with forty-six academics recognized as having experience or expertise in critical-thinking instruction, assessment, or theory. The experts reached an agreement on the core dimensions of critical thinking:

1. interpretation;
2. analysis;
3. evaluation;
4. inference;
5. explanation;
6. self-regulation.

The experts also reached consensus on the affective, dispositional components of critical thinking such as inquisitiveness, concern to become well-informed, and alertness to opportunities for critical thinking (Facione, 1990).

Halpern is also a noteworthy researcher in critical thinking. He describes critical thinking as purposeful, reasoned, and goal-directed. Halpern's approach to critical thinking places strong emphasis on decision-making, problem-solving, verbal reasoning, argument analysis, assessing likelihood and uncertainty, and hypothesis-testing (Halpern, 2003).

More recently, the Educational Testing Service (Liu, Frankel, & Roohr, 2014) proposed their own framework of critical thinking based on a review of existing critical-thinking frameworks. This framework consists of five dimensions:

- two analytical dimensions: evaluating evidence and its use and analyzing arguments;
- two synthesis-related dimensions: understanding implications and consequences and developing sound and valid arguments;
- one dimension relevant to both analysis and synthesis: understanding causation and explanation.

Looking across critical-thinking frameworks, there is agreement on the following components of critical thinking:

- looking at evidence and seeking justification;
- selecting pertinent information;
- distinguishing relevant from irrelevant facts;
- analyzing the credibility of an information source;
- determining the strength of an argument;
- identifying relationships and alternatives;
- discerning examples and counterexamples;
- recognizing assumptions, biases, and logical fallacies;
- defending ideas and hypotheses;
- drawing valid conclusions and inferences.

Across the different frameworks there seems to be less of an emphasis on skills related to understanding causation. Only two assessment-driven frameworks seem to emphasize hypothesis-testing or inferring causal relationships between variables (Halpern, 2010; Liu et al., 2014).

INSTITUTIONAL MODELS

Additionally, there are several disciplines that have their own critical-thinking frameworks. The American Institute of Certified Public Accountants Core Competency Framework (2017) gives the following skills related to critical thinking:

1. Decision modeling
   - Objectively consider issues
   - Identify alternatives
   - Implement solutions that provide value
2. Risk analysis
   - Identification and management of audit and business risk
3. Measurement
   - Uses various measurement and disclosure criteria used by accounting professionals and tax reporting
4. Research
   - Apply research skills to access relevant guidance or other information.

The American Psychological Association's Principles for Quality Undergraduate Education in Psychology include the following standard: "Faculty foster critical thinking by identifying the critical-thinking skills and abilities they wish to promote in their classes and in the psychology major as a whole. Faculty periodically review these skills and abilities throughout the term and through all years of undergraduate education" (APA, 2011, p. 9).

There are many foundations that have proposed frameworks around critical thinking. (For a review see Markle, Brenneman, Jackson, Burrus, & Robbins, 2013). None of these foundations offer assessments for their critical-thinking framework. This is due to the notion that critical-thinking assessment should be embedded in specific curriculum rather than be assessed devoid of course content. Many foundations do offer rubrics for guidance on critical-thinking assessments in disciplines. For example the American Association of Colleges and Universities initiated the Valid Assessment of Learning in Undergraduate Education (VALUE) initiative, an effort to establish rubrics for critical thinking in higher education. The VALUE rubrics have been adopted in some form at more than 5,600 institutions, including schools, higher-education associations, and more than 3,300 colleges and universities in the United States and around the world. Skills included in the VALUE framework refer to:

- clear explanation of issues;
- evidence construction and analysis;
- evaluation of context and assumptions;
- articulate multiple perspectives;
- logical conclusions.

Areas of Disagreement around Critical Thinking

Despite, and perhaps because of, the panoply of research around critical thinking, instructors still struggle with what critical thinking actually means. Bahr (2010) revealed that instructors teaching critical thinking do not have a clear understanding of what critical thinking means, with 37 percent of academics acknowledging the dispositional and self-regulatory aspects of critical thinking and only 47 percent describing critical thinking as involving processes or skills.

Another area of disagreement revolves around the role of background knowledge on critical thinking. 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. Others 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 (Bailin, Case, Coombs, & Daniels, 1999; Facione, 1990). Conversely, some researchers argue that critical-thinking skills can be generalized across different contexts and domains and can thus be taught in a generic way (e.g., Halpern, 2001; Lipman 1988; Van Gelder, 2005).

A final debated topic in critical thinking is around the role of dispositions (e.g., open-mindedness). Some researchers propose that dispositions or affective components play a role in critical thinking while others do not (e.g., Ennis, 1985; Facione, 1990). Empirical evidence appears to confirm the notion that critical-thinking abilities and dispositions are, in fact, separate skills (Facione, 2000). These dispositions have variously been cast as attitudes or habits of mind. Facione (2000, p. 64) defines critical-thinking dispositions as “consistent internal motivations to act toward or respond to persons, events, or circumstances in habitual, yet potentially malleable ways.” Researchers tend to identify similar sets of dispositions as relevant to critical thinking. For example, the most commonly cited critical-thinking dispositions include:

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

The next section will introduce a new model that attempts to resolve some of the debate around domain specificity and dispositions.

Pearson Critical-Thinking Framework

We propose a new framework to provide clarity on two broad areas for disagreement around critical thinking: the issue of domain specificity and the role of dispositions. First, we view critical thinking as a set of skills that can be defined in a general way and that have broad applicability across multiple disciplines, but which rely on subject-specific knowledge, conventions, and tools – intrinsic to a particular domain and discipline – for their expression. This explains why a person may demonstrate critical thinking in one domain and fail to do so in another.

Second, we deviate from previous approaches to defining critical thinking that have included dispositions. The writings of Socrates, Plato, Aristotle, Matthew Lipman, Peter Facione, and Richard Paul focus on the qualities and characteristics of a person rather than the behaviors or actions the critical thinker can perform (Facione, 1990; Lewis & Smith, 1993; Paul, 1992; Thayer-Bacon, 2000). One limitation of this approach to defining critical thinking is that it does not provide clear pedagogical value. That is, emphasizing traits or dispositions as part of critical thinking obfuscates the notion of critical thinking as a skill that that can be developed through instruction. Some evidence suggests that although having a disposition to think critically may increase the probability that a person does think critically in any given situation, it does not appear to have an effect on whether that person can learn to think critically. For example, Heijltjes, Van Gog, Leppink, and Paas (2014) found that students with high values on dispositions such as open-mindedness and enjoyment of effortful cognitive activity were no more (nor less) likely than those with low values of those dispositions to benefit from critical-thinking skills instruction. Thus, our framework does not include critical-thinking dispositions.

These two aspects of our framework distinguish it from previous work on critical thinking. The Pearson Critical-Thinking Framework (see Table 1) includes four main dimensions:

1. systems analysis;
2. argument analysis;
3. creation;
4. evaluation.

SKILL	DESCRIPTION	EXAMPLE BEHAVIORS	RELATED SKILLS
Systems analysis	Identifying and determining the relationships between variables to understand a system	Identify variables Test hypotheses Control for third variables	Troubleshooting Systems thinking Problem-solving Scientific reasoning Analysis Hypothesis-testing
Argument analysis	Drawing logical conclusions based on data or claims.	Identify claims to support a position Avoid cognitive biases (e.g., confirmation bias) Draw valid conclusions from a data analysis	Deduction Induction Problem-solving Reasoning Decision-making Data and information literacy Inference Data analysis
Creation	Creation of a strategy, theory, method, or argument based on a synthesis of evidence The artifact that is created goes beyond the information at hand	Provide arguments from multiple perspectives in a synthesis Develop a new tool to test compounds in a solution	Synthesis Computational thinking Dialectic debating Designing Planning
Evaluation	Judgment of the quality of procedures or solutions. Involves criticism or a work product using a set of standards or specific framework	Use an ethical framework to judge if a business is violating ethical principles in their accounting records Determine if an electrical installation in a home meets a standard of safety	Criticism Auditing Appraisal Authentication

Table 1 Pearson Critical-Thinking Framework with example behaviors and related skills.

SYSTEMS ANALYSIS

Systems analysis involves identifying variables within a system and determining the relationships between variables in a closed system. Systems analysis involves independent manipulation of variables, each associated with consequences in environments that may change dynamically (Funke, 2003). Whether it be troubleshooting a faulty modem or determining how a remote control works, systems analysis attempts to understand the relationship between variables in a closed system.

In many respects, systems analysis is similar to what researchers call problem-solving. The problem-solving literature is large but does not often delve deeply into specific cognitive processes used to solve problems, which makes it difficult to teach and assess. Troubleshooting, one type of problem-solving (Jonassen & Hung, 2006) does have more clearly defined elements, including:

- domain knowledge (general principles related to system);
- system or device knowledge (knowledge specifically about the device type);
- visual-spatial knowledge (memory of where parts are);
- procedural knowledge (routine rules on how to manipulate components in a system);
- strategic knowledge (how to perform tests and information-gathering activities);
- experiential knowledge (historical knowledge of past solutions).

For example, an expert IT mechanic is able to deeply understand a system as a function of

1. rich domain knowledge about principles of hardware and software;
2. familiarity with the target device in terms of the location of all of the components (motherboard, hard drive, RAM);
3. visual-spatial knowledge about where the hardware parts are;
4. procedural knowledge of flow states between, for example, RAM and the ROM and the reasons for changes in them;
5. strategic knowledge on how to test the hard drive in isolation;
6. experience with a variety of situations where a hard-drive failure is really a short circuit in the motherboard.

Thus, systems analysis is the process of problem-solving that seeks to understand how a system works.

ARGUMENT ANALYSIS

Argument analysis involves drawing conclusions based on evidence. Analysis of arguments involves breaking information down into their component parts, identifying claims and evidence, clarifying information needed to make a decision, and determining whether a conclusion is supported by evidence or logic (Facione, 1990). Argument analysis can also involve making conclusions based on data or data analysis. It is important to contrast analysis of arguments with construction of arguments, which we consider to be a part of the skill creation.

CREATION

Creation involves the production of a communication, proposed set of operations, or derivation of a set of abstract relations (Bloom et al., 1956). Specifically, the proposed set of operations can be an argument, method, procedure, algorithm, or formal mathematical equation. Creation comes from Anderson and Krathwohl's (2001) revision to Bloom's original taxonomy and corresponds to what was originally called synthesis. Accordingly, creation requires one to integrate disparate pieces into a new and coherent whole. The artifact that is created goes beyond the information at hand. Examples include creating an argument around global warming or designing an experiment to test the relationship between two variables. Creation requires either systems analysis or argumentation analysis.

For example, creation of arguments can involve resolution of differences of opinions, questions, and issues in critical discussions (Jonassen & Kim, 2010). From this perspective, learners build arguments to support a position, consider and weigh arguments, and clarify uncertainties in order to achieve a deeper understanding of complex issues (Cho & Jonassen, 2002).

EVALUATION

Evaluation is the judgment of the quality of procedures or solutions (Bloom et al., 1956). Evaluation is often a criticism of work products with respect to their credibility, relevance, and the potential for omissions or and bias (Facione, 1990). Evaluation involves progressing beyond merely identifying the source of propositions in an argument to actually examining the credibility of those identified sources using established standards (Bloom et al., 1956). Examples include reviewing an expense report according to accounting standards and ensuring a building permit followed safety codes (based on a standard). Evaluation requires either systems analysis or argumentation analysis.

Skills Related to Critical Thinking

There are many skills mentioned in the literature that can be seen as related to critical thinking. We review cognitive biases, computational thinking, systems thinking, and data literacy and describe how they are related to or accounted for in our framework.

AVOIDING COGNITIVE BIASES

Daniel Kahneman has studied cognitive biases extensively over his fifty-year career. Kahneman's 2011 book, Thinking, Fast and Slow, summarizes research that he conducted over the decades. The book highlights several decades of research suggesting that people place too much confidence in human judgment. Much of the book focuses on cognitive biases and how they influence our decision-making. For example, confirmation bias is the tendency to seek information that aligns with and confirms our preexisting beliefs. These types of biases are what critical thinking seeks to guard against; by improving argument analysis skills, for example, we hope to prevent people from unconditionally accepting the argument that agrees with their point of view.

Cognitive biases are related to the critical-thinking skill of argumentation analysis since they can impede a learner from reaching valid conclusions. The Pearson Critical-Thinking Framework does account for cognitive bias since this behavior could be a result of faulty logic in an argument.

COMPUTATIONAL THINKING

Computational thinking entails "solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science" (Wing, 2006, p. 33) and represented in a form that can be effectively carried out by an algorithm (Wing, 2011). The following skills are now widely accepted as comprising computational thinking and form the basis of curricula that aim to support its learning as well as assess its development (Grover & Pea, 2013):

- abstractions and pattern generalizations;
- systematic processing of information;

- symbol systems and representations;
- algorithmic notions of flow of control;
- structured problem decomposition (modularizing);
- iterative, recursive, and parallel thinking;
- conditional logic;
- efficiency and performance constraints;
- debugging and systematic error detection.

Computational thinking holds some similarities to the systems analysis critical-thinking skill since it focuses on efficiency in solutions and debugging and breaking down problems into smaller pieces. There is also an aspect of argument analysis as it relates to understanding rules of logic. There are also similarities to creation since it involves creation of a solution to a problem (in the form of code). However, the Pearson Critical-Thinking Framework does not include skills in computational thinking referring to domain knowledge around programming.

SYSTEMS THINKING

Systems thinking is defined as a holistic perspective that sees the world as increasingly interconnected, from elemental components to complex systems of activity (e.g., Sweeney & Sterman, 2007). Systems thinking involves deep understanding of the underlying social, natural, and technological relationships in a closed system.

Systems thinking holds many similarities to the critical-thinking skill of systems analysis in that they both focus on understanding systems as a whole not just the relation between particular parts. The Pearson Critical-Thinking Framework accounts for skills in systems thinking.

INFORMATION AND DATA LITERACY

Information literacy, is the ability to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (Association of College and Research Libraries, 2015). Other information-literacy standards include the Partnership for 21st Century Learning's Framework for 21st Century Learning (P21, 2015), which includes categories of information literacy, media literacy, and information and communications technology (ICT) literacy.

Calzada Prado and Marzal (2013) propose the following five data literacy standards:

1. understanding data;
2. finding and/or obtaining data;
3. reading, interpreting, and evaluating data;
4. managing data;
5. using data.

There is some overlap between the skills of information literacy and argument analysis, particularly encouraging learners to be skeptical of claims, to question the credibility of information sources, and to be aware of the ways in which an author's point of view can affect how information is communicated. Data literacy holds many similarities to argument analysis in that they both focus on building arguments with data. The Pearson Critical-Thinking Framework accounts for the skills in data and information literacy including locating data or information sources, managing data and information, and ethical citation of information sources.

Development of Critical Thinking

Given all of the definitions of critical thinking described here, it is perhaps not surprising that there is not a universally accepted description of progression of critical-thinking skills. Kuhn (1999) focused on the development of the ability to think about one's own and others' thoughts. She notes that somewhere between the ages of three and five, children begin to understand that what someone says is a statement of what they believe. That is, they develop the ability to understand that there is an internal universe of beliefs. Children also develop the idea that there can be false beliefs and that beliefs are not just mental copies of reality. However, even by eight years of age, children still struggle with the notion that two people could look at the same piece of evidence and reach different conclusions. This difficulty of thinking about the role of people in constructing their own knowledge persists through adolescence and sometimes into adulthood. They stay at what Kuhn terms an "absolutist" stance in which disagreements about claims can be resolved by going back and observing the evidence or appealing to authority. From here, people may end up believing that everyone has an opinion and that all opinions are equal, or they may progress to an "evaluative" stance in which claims can be weighed on the basis of their evidence in a process of judgment, evaluation, and argument. In this view, critical thinking is the method by which sound claims can be determined.

Looking at more specific skills, in systems analysis, a review of the literature conducted by Cheng, Ructtinger, Fugii, and Mislevy (2010) revealed a three-stage trajectory. At each stage, individuals developed further in three areas:

1. system characteristics;
2. reasoning about systems;
3. modeling systems.

For example, in the initial stage, individuals have difficulty recognizing that there may not be a singular causal force underlying a system. This progression posits that the initial stage is likely to last through middle school, on average. Across later stages they come to understand both that multiple variables can impact an outcome and also that a single variable can have multiple outcomes. It is proposed that the second stage is displayed throughout most of high school, with the third stage developing in the post-secondary period.

Similarly, in argumentation, researchers at the Educational Testing Service (Song, Dean, Graf, & Rijn, 2013) have proposed a five-level learning progression. At each level, they describe the development of four skills:

1. social appeal building;
2. argument building: taking a position;
3. argument building: reasons and evidence;
4. discourse: framing a case.

The first stage of the progression is thought to begin at the kindergarten level, with the drawing, dictating, and writing of pieces in which they express an opinion or preference. However, it is not until sixth grade that students move to writing arguments with relevant evidence, which is the second level in the progression. This progression of moving from expressing opinion to supporting an argument's claim with evidence is also consistent with the English Language Arts standards of the Common Core.

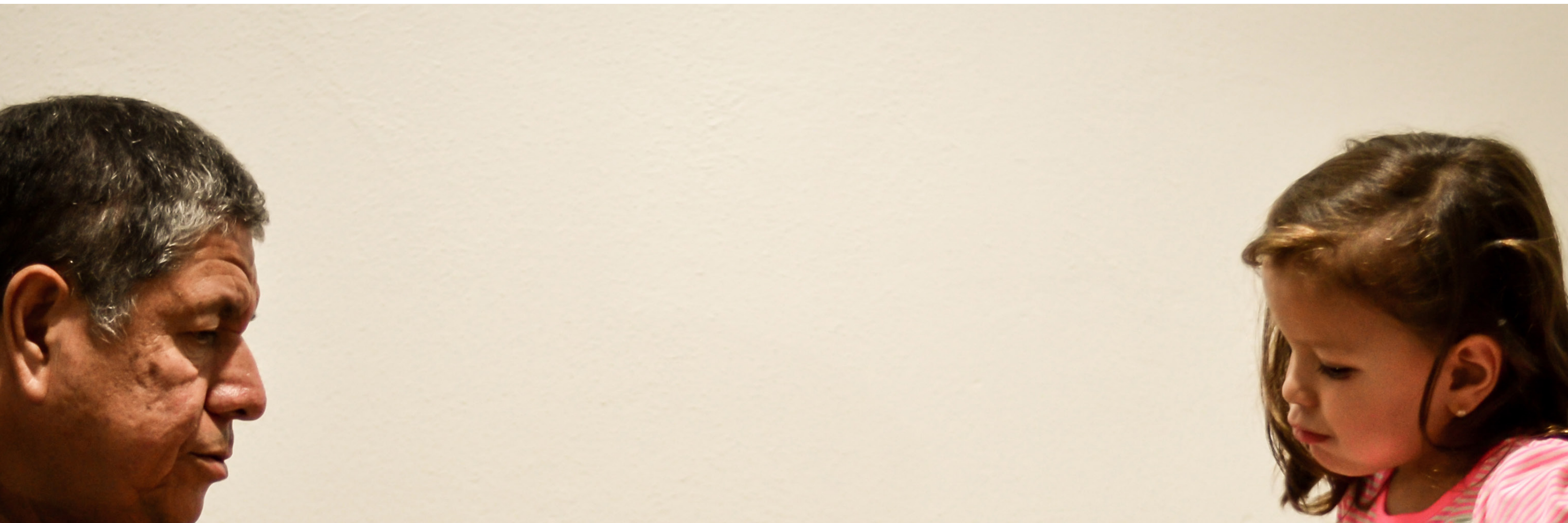

Image by Marcela Robles Villela

Teaching Critical Thinking

If we believe there are progressions of critical-thinking skills, we then want to understand how to move students along that progression. There are efforts to teach critical thinking generally in both K-12 and post-secondary environments. We will review research evaluating the effectiveness of those efforts and then look at efforts related to teaching specific subskills of critical thinking.

General Approaches to Teaching Critical Thinking

Many studies in both K-12 and higher-education settings compare the effects of different instructional models on the development of general critical-thinking skills, using Ennis's 1989 framework to categorize and contrast approaches:

- General: This approach to teaching critical thinking is usually typified by separate, stand-alone critical-thinking courses that are divorced from any particular discipline or domain and that explicitly teach critical-thinking principles outside of any particular subject-matter content.
- Infusion: This approach involves explicitly teaching principles of critical thinking embedded in specific subject-matter content within a discipline.
- Immersion: This approach does not explicitly teach critical-thinking skills; rather, there is an assumption that students will naturally develop critical thinking as a result of exposure to high-quality instruction in the discipline.
- Mixed: A hybrid approach, combining elements of both the general approach and either the infusion or immersion approach. Students receive explicit instruction in critical thinking as a separate thread of a larger course or program in the discipline or domain.

Abrami et al. (2008) conducted a meta-analysis of 117 studies of critical-thinking interventions targeting learners in K-12 and higher education as well as adult learners outside of school settings. The average effect size of these interventions was 0.341, although there was significant heterogeneity in effect sizes across studies. Follow-up analyses indicated that effect sizes were significantly larger for K-12 students than for undergraduates. In addition, mixed instructional approaches, in which critical-thinking skills are taught as a stand-alone topic or module within discipline- or subject-specific courses, were the most effective, whereas immersion approaches were the least effective. Moreover, interventions were more effective when instructors received extensive training to teach critical thinking and when frequent classroom observations of critical-thinking teaching practice were conducted. Finally, interventions employing collaborative learning approaches enjoyed a slight advantage over those that did not use group learning.

Teaching Critical Thinking in K-12

Many researchers have noted that critical-thinking skills are unlikely to develop in the absence of explicit instruction (Abrami et al., 2008, Facione, 1990, Halpern, 1998). The question then becomes: What should that instruction look like?

Published K-12 studies primarily use an infusion approach, specifically teaching critical-thinking skills in the context of a particular topic. A number of studies showed positive effects with this approach on critical thinking within that domain, but not on a more generalized measure of critical thinking. For example, students in a history class who received two weeks of instruction on generalization of ideas from historical events improved on their ability to generalize but showed no differences from the control group on a generalized measure of critical thinking. Similarly, a group of science researchers (Zohar & Tamir, 1993) found that students who completed specific activities to address ten elements of critical thinking in biology then performed better on an assessment of those elements than students who did not complete those activities. However, there was no difference between the groups on a general measure of critical thinking.

In a rare study of a general approach to teaching critical thinking to secondary-school students (Marin & Halpern, 2011), those who participated in a web-based critical-thinking workshop outperformed students who received training embedded in a class and students who did not receive training on a measure of general critical thinking. It is possible that these results are due to the fact that more generic critical-thinking instruction is more aligned to the more generic assessments of critical thinking, which results in the better performance.

Instruction in a particular domain is no guarantee of success in improving critical-thinking skills even within that domain. While the science study above found positive results, a separate study (Germann, 1989) used directed inquiry methods to teach experimental group students both science process skills and scientific problem-solving without measurable success. After an entire year of instruction in the curriculum there was no difference between the experimental group and the control group on science process skills or critical thinking. There are enough differences between the two studies that it is impossible to identify whether it is the student population, the instructional method and activities, or the outcome measures that resulted in competing findings. However, it is clear that focusing on teaching inquiry skills in science in and of itself is not sufficient to improve critical thinking.

SIMPLE INTERVENTIONS

Both of the science inquiry intervention studies above involved activities and combinations of teaching strategies. There is some evidence that smaller changes can improve skills. A qualitative study of a single elementary classroom conducted over two years indicated that as the classroom teacher moved questioning from factual recall to divergent questioning, students made more claims and supported those claims with strong evidence as well as refuting claims based on evidence (Martin & Hand, 2009).

SMALL-GROUP DISCUSSION

One technique with promise is using small-group discussions to teach critical thinking, particularly with elementary-aged children. (It is relatively untested in secondary classrooms.) In two separate studies by the same research group (Hudgins & Edelman, 1988; Riesenmy, Mitchell, Hudgins, & Ebel, 1991), one with fourth- and fifth-graders and the other with fifth- and sixth-graders, students in an experimental group engaged in eight to twelve small-group sessions where they learned and practiced “self-directed critical-thinking” skills. This involved both cognitive and motivational aspects of critical thinking. Students then individually completed problem-solving tasks. Students who had received the instruction used the skills more, used more information to solve the problem, and had higher-quality solutions than students who were not given instruction.

Separately, another research team sought to particularly address analysis of arguments (Klein, Olson, & Stanovich, 1997). Students in the experimental group read, wrote, and discussed arguments in small groups. The researchers also compared instruction on argument components with instruction on organizing argument elements. Students in all of the experimental groups subsequently performed better on an assessment of evaluating arguments and showed transfer of that skill from social-science to science problems.

Unfortunately, across the literature there are few other common instructional techniques that have shown to be effective in increasing critical thinking among primary and secondary students.

Higher Education

There is a large body of literature looking at the effects of the aggregate college experience on student cognitive development (e.g., Gellin, 2003, Loes, Pascarella, & Umbach, 2012; Pascarella & Terenzini, 2005). According to this literature, the college experience includes both the direct effects of instruction and social interactions with peers and faculty, as well as the indirect effects of institutional characteristics and the overall environment. Huber and Kuncel (2016) meta-analyzed seventy-one studies examining changes in overall critical thinking during the college years, concluding that after four years critical-thinking skills of students tend to improve by 0.59 standard deviations.

SPECIFIC INSTRUCTIONAL MODELS

There are also numerous studies documenting critical-thinking interventions carried out in college courses across a wide variety of disciplines. These studies often borrow Ennis's (1989) instructional models framework for comparing and contrasting different intervention types. For example, Behar-Horenstein and Niu (2011) reviewed forty-two empirical studies describing critical-thinking interventions, noting that a lower percentage of studies employing the immersion approach found significant gains in critical-thinking skills compared to the general, infusion, or mixed approaches. The authors also found mixed results for different instructional methods (e.g., concept mapping, problem-based learning), with no one instructional method showing consistently positive results. They did, however, conclude that, regardless of the instructional activities used, interventions lasting fewer than five months were unlikely to yield significant improvements in critical-thinking skills.

In a follow-up study, Niu, Behar-Horenstein, and Garvan (2013) meta-analyzed thirty-one quantitative studies examining the effect of instructional interventions on college students' critical-thinking skills. The overall effect size for such interventions was 0.195, with significant variability in effect sizes across studies. Effect sizes varied by discipline and duration, with interventions lasting fewer than twelve weeks less effective than those lasting longer than twelve weeks.

Rimiene (2002) describes a stand-alone, semester-length course focused on teaching critical thinking that featured an "active learning" pedagogical approach. The course explicitly taught critical-thinking principles and criteria and engaged students in problem-solving. Learning activities included brainstorming, reflexive writing, active listening, "purposeful research," class discussion, debates, and projects. Rimiene (2002) found that students randomly assigned to this course demonstrated significant gains in analysis, evaluation, inference, inductive reasoning, and deductive reasoning from pretest to posttest, whereas a control group that did not take the course, and thus did not receive explicit critical-thinking instruction, demonstrated no such improvements.

Problem-Based Learning

It is argued that efficient problem-solving is among the most important skills to learn (Jonassen, 1997). One commonly used approach to teaching problem-solving in the college classroom is problem-based learning (PBL). For example, Carriger (2016) explored the effectiveness of a PBL approach to teaching critical-thinking skills in the context of an undergraduate management course. In this study, principles of PBL instruction included:

- The instructor acts as a facilitator rather than a sage handing down wisdom.
- The process of problem-solving follows a carefully structured script.
- Real-world, ill-structured problems provide context.
- Learning is collaborative.
- Learning is assessed in relation to the explicitly stated course objectives.

Carriger (2016) compared the performance of students taught using a PBL approach to those taught either using a lecture-based approach or a hybrid approach (which combined shorter lectures with briefer ill-structured problems). Results suggested that students in both the PBL and the hybrid format demonstrated higher critical-thinking performance than did students in the traditional lecture format, with the hybrid format also fostering better knowledge acquisition of course concepts, particularly for low-achieving students.

Similarly, Zabit, Karagiannidou, and Zachariah (2016) compared the use of a PBL approach to a traditional, lecture-based approach in teaching critical-thinking skills to a group of economics undergraduates. Students were randomly assigned to either the PBL approach or to the traditional approach, and students' critical-thinking skills were assessed at the beginning, middle, and end of the semester. Although the researchers saw no significant differences in critical-thinking skills at the beginning or midpoint of the course, by the end of the semester, students exposed to the PBL approach significantly outperformed those in the traditional lecture condition, particularly with regards to inductive reasoning and analysis.

Şendağ and Odabaşı (2009) randomly assigned teacher candidates taking an online undergraduate course on technology-enhanced instruction to either a PBL approach or an instructor-centered approach to teaching. The PBL approach involved working through the following process with a set of ill-structured problems:

1. Introduce the problem.
2. Form groups.
3. Brainstorm prior knowledge and opinions on the problem.
4. Identify needed information to solve the problem.
5. Make a plan.
6. Execute the plan to solve the problem.
7. Evaluate team performance.

The instructor-centered approach did not use these ill-structured problems but used more traditional types of teaching activities, such as instructor-set discussion questions and student reports. In addition, in the PBL approach, the teacher acted as the facilitator and guide, whereas in the instructor-centered approach, the teacher acted as expert and activity director. Researchers found that students assigned to the PBL group exhibited significantly higher gains at posttest than did students assigned to the instructor-centered approach.

Kong, Qin, Zhou, Mou, and Gao (2014) meta-analyzed nine randomized controlled trials comparing the effectiveness of a PBL approach to traditional lectures in improving the critical-thinking skills of nursing students. In this study, PBL was defined as, “a student-centered approach to learning which enables the students to work cooperatively in small groups for seeking solutions to situations/ problems” (Kong et al., 2014, p. 459). The PBL process unfolded in five steps:

1. analysis of the problem or scenario;
2. establishment of learning objectives;
3. information collection;
4. information summarization;
5. reflection.

The authors reported an overall effect size of 0.33, suggesting that PBL approaches to teaching critical thinking are more effective than are traditional lecture-based approaches, although outcomes may depend on the length of the intervention (with longer interventions being more effective) and the type of outcome measure used.

Walker and Leary (2009) conducted a meta-analysis comparing PBL to traditional lecture across eighty-two studies spanning a variety of disciplines. Across all studies, the overall effect size was 0.13, with significant heterogeneity in effect sizes. The authors found that outcomes varied by discipline, with larger PBL effect sizes observed for teacher education and social-science interventions than for science and engineering interventions. Effects favoring the PBL approach were also more dramatic for interventions using more ill-defined problems and for studies in which the outcomes assessment targeted higher-level outcomes (application of concepts and principles to new contexts) rather than lower-level outcomes (defining or identifying concepts or articulating relationships between concepts).

Case Libraries

Experienced problem solvers in systems match new problems to prior experiences and apply those solutions (Aamodt & Plaza, 1994). Case libraries can serve as experiential knowledge and can be an effective form of instructional support for systems analysis (Jonassen & Hernandez-Serrano, 2002). The case library can consist of potentially hundreds of experienced problem-solvers’ solutions. Rather than relying only on a theoretical description of a system, the learner can access the case library to gain insight into systems (Jonassen & Hung, 2006).

Hernandez-Serrano and Jonassen (2003) investigated the effects of providing access to a case library of related stories to help solve ill-structured problems around food product development. The experimental group accessed experts’ stories of similar, previously solved problems, while a second group of comparable students accessed fact sheets (story content), and the control group accessed text selected at random from a textbook dealing with issues unrelated to the stories. Experimental students outperformed the comparable and control groups on multiple-choice questions assessing processes related to problem-solving (e.g., prediction, inferences, explanations). Performance on short-answer questions also assessing problem-related skills was not significantly different among the groups.

Worked Examples

Worked examples are instructional devices that typically model the process for solving a problem (Atkinson, Derry, Renkl, & Wortham, 2000). Worked examples can highlight the subgoals of the problem (Catrambone & Holyoak, 1990), which may include identifying fault symptoms, constructing a system model, diagnosing the fault, generating and verifying solutions, and adding experiences to the personal cases (Chi & Van Lehn, 1991). Schwonke et al. (2009) investigated worked examples in a geometry cognitive tutoring system. Students with access to worked examples needed less learning time to obtain the procedural skill and conceptual understanding of geometry than did students who did not have access to the worked examples. Van Gog, Kester and Paas (2011) investigated the effectiveness of worked examples for learning how to troubleshoot with electrical circuits. The results of the study showed that students in the worked-example condition significantly outperformed students in the traditional problem-solving conditions. Additionally, higher performance was reached with significantly lower effort during training (Van Gog et al., 2011).

Worked examples are not appropriate for all learners. Learners with high domain knowledge might find worked examples redundant and thus may lose interest in this form of instruction. This has been described as the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). Renkl, Atkinson, and Maier (2000) suggest that worked examples should be faded over time and replaced with problems for practice. Thus, it is important to consider the prior knowledge of the learner when using worked examples.

Concept Maps

The concept map is a widely used instrument for learning complex systems (Rittle-Johnson, Siegler, & Alibali, 2001). This tool can be described as a graphical illustration of a knowledge concept consisting of central terms that are represented as nodes linked by labeled arrows. The arrows represent the quality and the direction between the nodes (Alpert, 2006; Shemwell, Fu, Figueroa, Davis, & Shavelson, 2010). The process of constructing and evaluating concept maps for clarity, completeness, and accuracy is believed to support development of critical-thinking skills, and there is some empirical evidence to support this claim. For example, Yue, Zhang, Zhang, and Jin (2017) meta-analyzed eleven studies examining the effects of concept-mapping interventions on the critical-thinking skills of nursing students. The authors found that students who were taught to construct and use concept maps had significantly higher critical-thinking scores than did students in the control group, who were typically taught using more traditional, lecture-based approaches.

Teaching Specific Critical-Thinking Skills

ARGUMENT ANALYSIS

Argument analysis involves induction and deduction without creation of arguments. Thus, this might require taking an argument apart or judging whether a particular claim is supported but stops short of actually constructing arguments. Research suggests that both explicit approaches to teaching critical-thinking and collaborative learning arrangements may support development of argument-analysis skills.

Explicit Instruction

Penningroth, Despain, and Gray (2007, p. 153) describe a one-credit course specifically designed to improve students' psychological critical thinking, defined as "the ability to evaluate claims (primarily for psychological issues) by applying scientific concepts," such as experimenter bias, correlation versus causation, and the notion of confounding variables. The authors used an active learning approach that included several types of learning activities including small-group activities, class discussions in small sections, frequent quizzes, and a final writing assignment (which involved comparing the results from several studies to evaluate a general claim). Working in groups, students were required to evaluate a set of findings from a behavioral research study to identify the methodological flaw, identify potential confounds, or suggest improvements to the research design. In the final writing assignment, students had to evaluate the veracity of a given psychological claim based on the quality of the supporting evidence. The authors found that although both students in the critical-thinking course and students in a comparable psychology course with very little explicit critical-thinking instruction demonstrated significantly higher psychological argument analysis scores at posttest compared to pretest, the effect size for the treatment group was larger, and they significantly outperformed their control group counterparts in identifying flaws in a set of claims linked to evidence.

Bensley and Spero (2014) implemented what they called a “direct infusion” approach to teaching critical thinking, which included explicitly teaching critical-thinking principles embedded within subject-matter knowledge in a cognitive psychology course through the use of examples, guided instruction, direct instruction, and formative assessment and feedback on critical-thinking performance. The authors used several critical reading and argumentation exercises that required students to complete a short reading from the textbook on some aspect of cognitive psychology (e.g., whether flashbulb memory is accurate), to analyze and evaluate the strength of the evidence on this issue, and to draw an inductive conclusion on the basis of that evidence. Results suggest that the direct infusion group made significantly greater gains in argument-analysis skills (recognizing kinds of evidence, evaluating evidence, distinguishing arguments from non-arguments, identifying assumptions, drawing appropriate inductive conclusions from research) compared to a control group receiving direct infusion of instruction on memory improvement and another group who received traditional instruction focused on content knowledge acquisition.

In a study designed to explore how students can be trained to avoid biased reasoning in drawing conclusions, Heijltjes et al. (2014) tested the effectiveness of explicit instruction in biased reasoning and fallacies. The authors found that all three groups receiving explicit instruction significantly outperformed students in the control group on both an immediate critical-thinking posttest as well as a delayed critical-thinking posttest administered three weeks later, but only on the types of probabilistic reasoning tasks that had been practiced. There was no difference in the treatment and control-group performance on other reasoning task types that had not been practiced (e.g., syllogistic reasoning, framing). The authors conclude that practice of critical-thinking skills is necessary in order for explicit critical-thinking instruction to yield learning gains. Interestingly, students never received feedback on their practice-task performance, which suggests that the combination of explicit instruction and practice can be effective for analysis of arguments, even without feedback.

Collaborative Learning

Several studies have examined links between cooperative or collaborative learning environments and general critical thinking or problem-solving in college students (e.g., Cabrera, Colbeck, & Terenzini, 2001; Quitadamo, Brahler, & Crouch, 2009). Recently, Loes and Pascarella (2017) administered a pretest and posttest measure of critical thinking to students at the beginning and end of their first year in college across 1,455 students from across nineteen different US institutions. Students were asked to indicate the extent to which they were exposed to collaborative learning environments and found that for low-achieving white students participation in collaborative learning significantly improved their argument-analysis skills relative to students who did not have collaborative learning experiences.

SYSTEMS ANALYSIS

Systems analysis involves understanding the relationship between variables in a closed system. Research suggests that several approaches to teaching systems analysis, such as procedural approaches (providing a sequence of actions), conceptual approaches (emphasizing a theoretical understanding of the system), and rule-based approaches (following a set of rules for solving problems), are not particularly effective (see Jonassen & Hung, 2006). However, the use of simulations may be a promising approach.

Simulations

Simulations can be defined as a technology modeling a system or a process where a user can manipulate parameters in the system (De Jong & Van Joolingen, 1998). Processes, systems, or functions of real-life phenomena are simulated in an authentic manner to enable understanding of a system or device (Mayer, Dale, Fraccastoro, & Moss, 2011; Siewiorek, Gegenfurtner, Lainema, Saarinen, & Lehtinen, 2013). Johnson and Rouse (2001) found that practice on computer simulations resulted in learning that was comparable to that achieved through traditional lecture. And higher learning gains are achieved when simulations demonstrate high fidelity, or a high degree of similarity, to the physical systems they are designed to represent (Allen, Hayes, & Buffardi, 2001; Johnson & Norton, 1992). For example, Park and Gittelman (1992) found that an animated simulator teaching electronics troubleshooting resulted in shorter learning times and fewer trials than a static simulator. Lovelace, Eggers, and Dyck (2014) found that exposing business undergraduates to two different team-based simulations, both representing complex business systems, demonstrated significant gains in their ability to analyze a business case and to make a clear recommendation regarding the appropriate course of action, taking into consideration the implications and consequences of those actions within hypothetical business contexts.

CREATION

Creation involves construction of an artifact by synthesizing and integrating disparate pieces into a new, coherent whole.

Argumentation

Most of the research on teaching argument creation falls under formal argumentation, where learners construct arguments and exchange them in dialogues to resolve different standpoints on the issues to find solutions for complex problems (Baker, 2003; Kollar et al., 2007; Stegmann et al., 2012). The dialectical form of argumentation – arguing both sides of an issue – is often used (Noroozi, Weinberger, Biemans, Mulder, & Chizari, 2012) to enable learners to perceive multiple perspectives in an argument (Jonassen & Kim, 2010). Dialectical argumentation in practice refers to expressing all possible evidence to support opposing claims, which are then clarified, contested, and refined through critical dialogue (Ravenscroft, 2011).

Exposure to dialectic argumentation has been argued to contribute to the skill of argument construction (Kuhn, Hemberger, & Khait, 2015; McAllister, Ravenscroft, & Scanlon, 2004). Supporting this notion, O'Keefe (1998) ran a meta-analysis of 107 studies on argumentation, revealing that reading dialectic arguments resulted in creation of more persuasive arguments than reading non-dialectic arguments. Additionally, Nussbaum and Kardash (2005), found that by including a counterarguments prompt, students significantly increased the number of counterarguments and rebuttals compared to those students who received no such prompt. Nussbaum and Schraw (2007) created two treatments to test argumentation instruction:

1. a graphical organizer on how to structure an argumentative essay;
2. an explicit instruction in argumentative writing.

Both treatments had a positive effect on production of counterarguments and rebuttals. Specifically, the explicit instruction was found to be more effective in counterargumentation production than the graphic organizer.

Computer-supported collaboration scripts (CSCLs) support collaborative argumentation and argumentative knowledge construction in digital environments (see Noroozi et al., 2012, for a review of features of CSCLs). CSCLs provide detailed and explicit guidelines for small groups of learners around argumentative knowledge construction (Weinberger, Stegmann, & Fischer, 2007). Prompts are also given to provide learners with guidelines, hints, and suggestions to facilitate effective argumentation strategies (Ge & Land, 2004; Noroozi et al., 2012; McAllister et al., 2004). In a CSCL interface, the student is prompted to enter a claim, rationale for the claim, and limitations of the claim. This information is then shared with other students via a discussion board to facilitate collaborative argumentation. There is empirical evidence that CSCL can improve the construction of counterarguments (Nussbaum, Hartley, Sinatra, Reynolds, & Bendixen, 2004) and sound arguments (Yiong-Hwee & Churchill, 2007), quality of argumentation in transfer tasks after training (Jermann & Dillenbourg, 2003; Stegmann, Weinberger, & Fischer, 2007; Schellens, Van Keer, De Wever, & Valcke, 2007; Weinberger et al., 2007), and co-construction of arguments with peers (e.g., Rummel, Spada, & Hauser, 2009; Schellens et al., 2007; Weinberger et al., 2007).

Other researchers have shown explicit writing instruction around identifying claims and evidence in writing can improve critical thinking. Bensley and Haynes (1995) provided students with a “thinking frame” for structuring their argumentative thinking and writing. This template required students to

1. identify the claim used in an argument;
2. evaluate the evidence supporting or refuting the claim;
3. draw and substantiate a conclusion about the truth of the claim.

The authors found that students trained to use these templates performed significantly better on a persuasive writing assignment at posttest than students not using the template. Angeli and Valanides (2009) investigated writing strategies such as analyzing the problem, generating solutions, developing the reasoning for each solution, deciding which is the best solution, and using criteria to evaluate one’s thinking. Results demonstrated that students assigned to the writing strategy significantly outperformed those in the control group on their ability to construct a coherent written argument.

Finally, Blessing and Blessing (2010) describe an Intro to Psych assignment called “PsychBusters,” modeled after the popular television show MythBusters, which requires students to work in teams of three to research a specific psychological myth or misconception (e.g., “listening to Mozart makes you smarter”), explain the basis for the myth, evaluate evidence supporting or refuting the myth, and make a decision as to whether the myth is “confirmed,” “plausible,” or “busted.” The authors found that students randomly assigned to a course using this assignment significantly improved their ability to create an argument from pretest to posttest, whereas students in the other course showed no such improvement.

Summary of Teaching Critical Thinking

Critical-thinking instruction has been studied extensively over the past twenty years, resulting in several meta-analyses comparing a variety of critical-thinking instructional approaches (e.g., PBL, immersion). The focus of much of this research has been primarily on argument analysis and argument creation showing favorable results of instruction on persuasive writing and avoiding fallacies. Research is growing around teaching systems analysis as a result of advances in technology around simulations.

Image by Christo VanDerWalt

Assessing Critical Thinking

Learning often involves deliberate practice through repeated interactions with learning activities (Ericsson, 2016). Similarly, learning how to think critically involves practice. In order to maximize learning, we need assessments that properly elicit or require critical thinking. In this section, we review published measures of domain-general critical-thinking skills. Then we explore ways of thinking about designing and scoring critical-thinking tasks within a particular discipline.

Domain-General Measures

There are a variety of published measures that seek to assess generalized critical-thinking skills. These include:

- California Critical Thinking Skills Test (CCTST);
- Halpern Critical Thinking Assessment (HCTA);
- Watson–Glaser Critical Thinking Appraisal (WGCTA; Watson & Glaser, 1980);
- Ennis–Weir Critical Thinking Essay Test (Ennis & Weir, 1985);
- Cornell Critical Thinking Test (CCTT; Ennis, Millman, & Tomko, 1985);
- ETS Proficiency Profile (EPP; ETS, 2010);
- Collegiate Learning Assessment+ (CLA+; Zahner, 2013);
- Collegiate Assessment of Academic Proficiency (CAAP Program Management, 2012).

The CAAP, CCTST, and WGCTA exclusively use selected-response items such as multiple-choice or Likert-type items, while the EPP, HCTA, and CLA+ use a combination of multiple-choice and constructed-response items. The Ennis–Weir test is an essay-only test.

Kobrin, Sato, Lai, and Weegar (2016) conducted an alignment study comparing the constructs assessed by eight published critical-thinking assessments to a definition of critical thinking that included cognitive, metacognitive, and dispositional components. The authors concluded that the published measures appear to be adequately aligned to the cognitive components, particularly identifying central issues and assumptions, but are rather poorly aligned to both the dispositional and metacognitive components. In terms of performance complexity, nearly all the items reviewed on these assessments required learners to process and use information in order to render a judgment about whether a particular inference or claim was supported. However, owing to limitations in the item-types used (five of the eight assessments relied exclusively on selected-response formats), few items required learners to generate their own inferences and conclusions. Thus, published measures of general critical-thinking skills may assess only limited aspects of what it means to think critically.

At the same time, published measures may be tapping general cognitive ability rather than critical thinking. Liu et al. (2014) reviewed some of the more recent validity studies for published tests of critical thinking, which show positive moderate correlations with general standardized tests, course grades, GPA, and job performance and negative correlations with reported negative life events. However, critics of these general measures argue that they may not be measuring distinct skills from cognitive ability

Critical Thinking in Practice

Filtering out biased or misleading information

Anita Woolfolk has been teaching educational psychology for over forty years. She has written several textbooks on educational psychology placing emphasis on critical thinking. One challenge in teaching educational psychology is demonstrating the relevance of the content to actual teaching practice. In her writing and teaching she has addressed this challenge using "What Would You Do?" case studies. Every chapter of her Educational Psychology (forthcoming) text begins with a case asking the readers, "What Would You Do?" followed by several critical-thinking questions. She begins with this case:

> *Imagine you are a teacher and assigned a research paper. As students start to ask questions you find more and more students are using the Internet for all their information. In itself, using the Internet is not bad, but the students appear to be completely uncritical about what they find on the Internet. "If it is on the web, it must be right" is the attitude of most students. Their first drafts are filled with quotes that seem very biased, with no sources cited or listed. It is not just that students don't know how to reference their work. You are more concerned that they cannot critically evaluate what they are reading. And all they are reading is the Internet!*

Assessments

How would you help your students evaluate the information they are finding on the Internet?

- Beyond this immediate issue, how will you help students think more critically about the subjects you are teaching?
- How will you take into account the cultural beliefs and values of your students as you support their critical thinking?

(Liu et al., 2014; Kuncel, 2011). For example, Liu et al. (2014) point out that because some of these studies do not control for potential confounds, it may be overall cognitive ability rather than critical-thinking skills that is driving these relationships.

There are also psychometric limitations to published measures. Because of the multifaceted nature of critical thinking, many published assessments include multiple subscales. The advantage of subscale scores is that they can provide more diagnostic information about specific skills. However, subscale scores are sometimes highly correlated and therefore difficult to distinguish. For example, Bernard et al. (2008) conducted a meta-analysis on the WGCTA and found that the subscale scores (inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments) were difficult to statistically differentiate. In addition, Leppa (1997) found that the subscale scores of the CCTST (analysis, evaluation, inference, deduction, induction) had low internal consistency. And although published measures that offer at least some constructed-response activities have the potential to improve the authenticity of the assessment compared to those that rely exclusively on selected response, the downside of these activities is that they tend to have lower levels of reliability than selected-response items (Liu et al. 2014; Lee, Liu, & Linn, 2011; Zahner, 2013).

Liu et al. (2014) recommend that future research should clarify the relationship between general, standardized measures of critical thinking and life outcomes (e.g., job success) beyond what is predicted by other cognitive assessments. However, we think it is equally if not more important to study how to design tasks that provide consistent results and that can be validly used to provide formative feedback to students about their critical-thinking skills within specific disciplines. Published measures may have a role to play as summative outcome measures – a means for demonstrating whether critical-thinking interventions have been successful. But we argue that embedding well-designed, discipline-specific, formative critical-thinking assessments into these interventions is crucial if we are to improve upon the effectiveness of critical-thinking instruction.

Evidence-Centered Design

Despite the availability of a wide variety of published critical-thinking measures, educators may want to design their own assessments of critical-thinking skills. Such homegrown assessments can be better tied to learning objectives and subject matter than can published assessments of general critical thinking, providing both a closer match to the specific aspects of critical-thinking instructors want to target and a better measure of critical-thinking skills as they are practiced in a given discipline. Evidence-centered design provides a systematic framework for developing assessment tasks to elicit targeted skills (Mislevy, Steinberg, & Almond, 2003).

The evidence-centered design (ECD) framework consists of three models:

1. Student model: Define the claims to be made about learners' competencies.
2. Evidence model: Establish what constitutes valid evidence of the claim.
3. Task model: Determine the nature and form of tasks that will elicit that evidence.

Therefore, a good task for critical thinking elicits behavior that provides evidence about key critical-thinking skills, and it must also provide principled interpretations of that evidence in terms that suit the purpose of the assessment. In short, the ECD approach provides a framework for developing assessment tasks that are explicitly linked to claims about skills via an evidentiary chain. In the sections that follow, we describe and provide examples of task model types and evidence model types commonly used to assess critical thinking.

In the context of critical thinking, the student model is typically the skills outlined in a particular framework (in the case of this paper: systems analysis, argument analysis, creation, and evaluation as described above). We will describe task and evidence models aligned to these skills in more detail below.

Assessment Task Models

Task models describe types of activities students can engage in that are likely to elicit evidence of critical thinking. Below we describe a number of task types and discuss the elements of critical thinking they have been used to assess.

You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech's sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.

WRITING TASKS

Writing is a multifaceted ability that requires language, organization and expression skills as well as domain knowledge (Foltz, 2015). In general, writing tasks are an effective way to elicit analysis, creation, and evaluation skills. It is a widely accepted view that writing can be a valid way to show proficiency in argumentation (Graham & Perin, 2007; Kuhn, 1999; Kuhn, Hemberger, & Khait, 2015).

One example of a writing task used to assess critical thinking can be found in the CLA+. In the CLA+, respondents are given evidence (e.g., documents) that must be critically synthesized and organized into supporting claims. Students must identify a course of action to resolve conflicting or competing viewpoints and provide an argument for their decision. Tasks allow different conclusions to be supported. Scores in the CLA are assigned to responses to individual prompts, as well as more holistically. Holistic dimensions include:

- use of relevant information;
- recognizing strengths and weaknesses of information;
- constructing arguments;
- acknowledging alternatives;
- writing quality.

Below is an example task from the CLA (retrieved from Arum & Roksa, 2011).

You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech's sales force, recommended that DynaTech buy a small private plane (a

Critical Thinking in Practice

Put Yourself in Their Place

Another challenge noted by Woolfolk in teaching educational psychology is demonstrating the relevance of the content to actual teaching practice and to the lives of the readers. In her writing and teaching she has addressed this by asking students to "Put Yourself in Their Place." From Woolfolk's chapter on diversity:

> *Legal segregation came to an end in 1954. Take a moment to imagine you were living back then and the child described here was your own.*
>
> *[In] the city of Topeka, Kansas, a minister walked hand in hand with his seven-year old daughter to an elementary school four blocks from their home. Linda Brown wanted to enroll in the 2nd grade, but the school refused to admit her. Instead, public school officials required her to attend another school two miles away. This meant that she had to walk six blocks to a bus stop, where she sometimes waited half an hour for the bus. In bad weather, Linda Brown would be soaking wet by the time the bus came; one day she became so cold at the bus stop that she walked back home. Why, she asked her parents, could she not attend the school only four blocks away?*
>
> (Macionis, 2003, p. 353)
>
> *Her parents' answer to this question, with the help of other concerned families, was*

SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.

Students are provided with the following resources:

- newspaper articles about the accident;
- a federal accident report on in-flight breakups in single-engine planes;
- Pat Williams' email to her assistant and Sally Evans' email to Pat Williams;
- charts on SwiftAir's performance characteristics;
- an article from Amateur Pilot magazine comparing the SwiftAir 235 to similar planes;
- pictures and descriptions of SwiftAir models 180 and 235.

Students are then instructed to prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken into account, and your overall recommendation about whether or not DynaTech should purchase the plane.

SIMULATION TASKS

A simulation-based assessment (SBA) is the use of a simulation for purposes of assessment. In general, SBAs are good candidates for eliciting systems analysis. Specifically, SBAs are simulations with specific problems that must be solved in the simulation environment (Behrens, DiCerbo, & Ferrara, 2012; DiCerbo & Behrens, 2012; Gegenfurtner, Quesada-Pallarès, & Knogler, 2014). In the realm of systems-related SBAs, researchers have analyzed student performance in simulations in computer networking (e.g., West et al., 2010, Rupp, Nugent, & Nelson, 2012; Levy, 2013), architecture (Braun, Bejar, & Williamson, 2006), flood and fire damage control (Koenig, Lee, Iseli, & Wainess, 2010), medical diagnosis (Consorti, Mancuso, Nocioni & Piccolo, 2012; Margolis, & Clauser, 2006), decision-making in businesses (Siewiorek et al., 2010), dental practice (Mislevy, Steinberg, Almond, Breyer, & Johnson, 2001), urban planning (Rupp, Gushta, Mislevy, & Shaffer, 2010), and complex spatial reasoning (Ventura, Shute, Wright, & Zhao, 2013). While SBA is still a fairly new field, the research thus far supports its validity as an assessment approach. For example, Ventura et al. (2013) found that an SBA measuring large-scale spatial skills predicted academic performance in a STEM field better than an established pencil-and-paper assessment of the same skill.

to file a suit challenging the school policy. You know the outcome of the 1954 Brown v. The Board of Education of Topeka ruling. "Separate but equal" schools for Black children were declared inherently unequal.

Assessments

- What would you do in this situation?

DiCerbo and Behrens (2012) discuss several notable advantages of SBAs over other types of assessment activities (e.g., multiple-choice questions). First, SBAs provide a much larger range of activities that can be used to elicit a large range of responses. More specifically, SBAs can offer the possibility to capture students' process data as well as their product data. Product data can be regarded as the final work products that students produce during the simulation, whereas process data are log-file entries that indicate how students produced their work products (Rupp et al., 2012). Second, as a function of a larger amount of data collected on any given student, SBAs can produce more accurate estimates of key competencies. SBAs may also enjoy advantages over traditional paper-and-pencil assessments with respect to student engagement (De Klerk, Veldkamp, & Eggen, 2011; Gegenfurtner et al., 2014; Ventura et al., 2013). Third, by improving the inferences we can make about a student, we can also improve the feedback we can provide on how to fix misconceptions or fill gaps in knowledge. Thus, the use of process data means that a student's proficiency development can be measured throughout the learning experience (DiCerbo & Behrens, 2012; Shute & Ventura, 2013).

CONCEPT MAP TASKS

Concept maps represent the structure of knowledge in a domain (Erdogan, 2009; Schaal, Bogner, & Girwidz, 2010). In general, concept-mapping tasks are good at eliciting systems analysis and creation. These tasks allow students to illustrate complex relationships between concepts, which is difficult to do using traditional item types (Weinerth, Koenig, Brunner, & Martin, 2014). They can be utilized to assess complex knowledge structures in a variety of different subject areas (e.g., biology, chemistry, physics [Steiner, Albert, & Heller, 2007; Stoyanov & Kommers, 2008]). Moreover, concept maps have been shown to be particularly useful in diagnosing misconceptions (McClure, Sonak, & Suen, 1999). Concept maps are widely accepted as useful assessment tools in higher education (for a review, see Weinerth et al., 2014). There are two approaches to designing concept-map tasks: high-directed, in which concepts and links are provided and students must simply slot entities and links into the correct spot on the map; and low-directed, where students must generate both the concepts and the links themselves (Ruiz-Primo, 2000).

Concept maps have traditionally been implemented as paper-and-pencil assessments. Computer-based concept maps offer several advantages in comparison with paper-and-pencil-generated concept maps (Weinerth et al., 2014):

- Students can more easily construct, modify, or maintain concept maps.
- Software can deliver real-time feedback around student-constructed maps.
- Teachers can provide students with constraints around concept-map construction.
- Concept maps can be scored automatically and objectively.

Problem Types for Critical Thinking

Whereas writing tasks, SBAs, and concept maps describe different broad approaches to eliciting some of the same critical-thinking skills, other researchers have conceptualized different problem types that vary according to their cognitive demands. For example, Jonassen (2011) provides a typology of problems that can be used to guide assessment and instruction of critical thinking in specific disciplines. In Jonassen's framework, the skills of critical thinking can be seen as a continuum of cognitive processes that are elicited to varying degrees by different problem types (see Table 2).

PROBLEM	DESCRIPTION	CRITICAL THINKING SKILL ELICITED
Story	Require the learner to engage in a process of breaking up a story into relevant parts Generate the appropriate solution strategy and apply the solution strategy to generate an answer	Systems analysis Creation
Rule	Tend to have a clear purpose or goal that is constrained but not restricted to a specific procedure or method Require that learners induce the rules governing how some system operates	Systems analysis
Decision-making	Entail selecting from a set of alternatives and their associated consequences Involve associated activities, such as generating additional alternatives and assessing the risks and benefits of alternatives	Argument analysis Creation
Troubleshooting	Require resolving goal state and current state discrepancies Involves error detection in other contexts such as detecting errors in a written argument, mathematical calculation, or software code	Systems analysis
Strategic performance	Entail real-time, complex activities Learners apply a number of tactical actions aimed at solving an ill-structured problem Usually under time pressure	Systems analysis
Policy	Multiple positions and perspectives exist Include foreign policy, legal issues, and economic and development issues	Argument analysis Systems analysis Creation Evaluation
Design	Have vaguely defined or unclear goals with unstated constraints Applying a great deal of domain knowledge Possess multiple solutions, with multiple solution paths Possess multiple criteria for evaluating solutions, and these criteria are often unknown	Argument analysis Systems analysis Creation Evaluation
Dilemmas	Considered the most ill-structured kind of problem because there are typically no widely accepted solutions Many important perspectives to consider: constitutional, political, social, ethical	Argument analysis Systems analysis Creation Evaluation

Table 2 Problem types and critical-thinking skills elicited.

Jonassen's problem types can be crossed with the different assessment approaches in any given assessment. For example, an SBA in computer science may feature troubleshooting problems to assess systems-analysis skills, but a business simulation might employ decision-making problems to assess argument-analysis skills.

Evidence Models

The final piece in the assessment puzzle is identifying and aggregating evidence from these activities so we can make inferences about critical thinking. An evidence model (e.g., Mislevy, Steinberg, & Almond, 2003) describes the specific types of behaviors that should be measured and how those behaviors link to skills. Gathering evidence of critical-thinking skills, however, is often not as straightforward as gathering evidence of content mastery. Many of the behavioral variables needed to assess the skill are not typically captured by traditional standardized item formats (e.g., multiple-choice). These variables include testing relationships between tools in simulations and writing arguments. Thus, the evidence model for critical thinking must specify how to identify the behaviors and how they tie to the constructs that need to be measured.

In the sections that follow, we first discuss general types of evidence gleaned from writing tasks and for the Jonassen problem types. Then we present two specific evidence models: one for a troubleshooting simulation and the other for an argumentation game.

EVIDENCE FOR WRITING TASKS

Writing prompts tend to be scored using some sort of rubric that either provides an overall score to represent the quality of the response (holistic) or provides multiple scores for different traits of the response (analytic). Rubrics may also identify specific features of the written response that should be considered in assigning scores, and these features may vary depending on what aspects of critical thinking are the focus of the assessment. For example, if the purpose of the writing task is to measure a student's argument-creation skills, pertinent features might include whether the student explicitly made a claim, whether the claim was supported by relevant evidence, whether the student considered counterarguments, and whether the student evaluated the strength of evidence supporting counterarguments. Below is a rubric used to score students' responses to the prompt, "Should drugs be legalized?" Students were required to create a written outline of a response to that question, and researchers assigned an overall holistic score to the outline according to the criteria in the rubric (Angeli & Valanides, 2009).

Critical-thinking performance scoring rubric

1. Reason clearly within a point of view.
2. Discuss the issue from different perspectives.
3. Identify pros and cons for perspective.
4. Explain with reasons and evidence which perspective they think is best.

Human scoring of student writing is time-consuming and expensive to do with any high degree of consistency. In writing assignments, raters or instructors must carefully read essays and assign one or more scores. For formative assessment, they may also need to provide detailed feedback to students. Both these requirements make it difficult to use writing assignments at scale, either in large classes or in a large-scale assessment. Fortunately, new advancements in machine learning are enabling the automated assessment of writing. Automated scoring of writing, or Automated Essay Scoring (AES), provides the ability to analyze student writing and to assign a score instantly. AES has become increasingly accepted for scoring of writing (e.g., Shermis & Burstein, 2013). Studies of AES systems have shown that they can be as accurate as human scorers (e.g., Shermis & Hamner, 2012), can score on multiple traits of writing (e.g., Foltz, 2015), and can be used for feedback on content (Foltz, Gilliam, & Kendall, 2000). AES is now being used in a number of high-stakes assessments (e.g., various state K-12 assessments), placement tests (e.g., Pearson Test of English, ACCUPLACER), and for writing practice (e.g., Criterion, WriteToLearn).

AES has been used to accurately score a range of argumentation skills including:

- overall writing quality;
- content and ideas;
- focus and organization;
- development and details;
- conventions;
- coherence;
- reading comprehension;
- progression of ideas;
- style;
- appropriate examples;
- reasons;
- other evidence to support a position.

(Foltz, 2015)

EVIDENCE FOR JONASSEN PROBLEM TYPES

Table 3 provides example evidence for an example task for each of the Jonassen problem types. The evidence is delineated by the specific critical-thinking subskill being assessed. There can be multiple points of evidence for each skill within a given task.

PROBLEM TYPE	EXAMPLE TASK	SKILL ELICITED	EXAMPLE EVIDENCE TO SUPPORT SKILL
Story	Construct an equation to solve a practical problem	Systems analysis	Breaks the problem down into known and unknown variables and quantities
		Creation	Generates an equation to represent the problem
Rule	Determining how a remote works with a TV	Systems analysis	Applies a systematic procedure for testing the relationship between buttons on the remote and resulting consequences on the TV.
Decision-making	Which of two different investment strategies will yield higher earnings?	Argument analysis	Reviews the pros and cons for each investment strategy
		Creation	Generates an argument as to why one strategy is better than the other

PROBLEM TYPE	EXAMPLE TASK	SKILL ELICITED	EXAMPLE EVIDENCE TO SUPPORT SKILL
Troubleshooting	Determine why a car won't start	Systems analysis	Produces an accurate and complete drawing of components in a system and relationships between components to demonstrate the source of the problem
Strategic performance	Using a flight simulator	Systems analysis	Determines the correct combination of actions to avoid crashing when a right engine fails
Policy	What is the best way to lower taxes?	Argument analysis	Identifies all the ways to lower taxes
		Creation	Proposes a solution to lower taxes with supporting evidence
		Evaluation	Evaluates that solution with respect to its implications for overall budget
Design	Design a new way to reduce traffic	Systems analysis	Tests the relationship between different traffic solutions on traffic
		Creation	Creates a new layout for a set of traffic lights that will reduce traffic
		Evaluation	Evaluates that layout against the old one in terms of traffic efficiency
Dilemmas	Should the government provide universal healthcare?	Argument analysis	Assess the logical strength of arguments for and against universal healthcare
		Creation	Composes an argument supporting a position on the issue
		Evaluation	Evaluates a specific healthcare proposal in terms of net impact on society

Table 3 Problem types and evidence examples.

ASSESSMENT EXAMPLES

Evidence Model for a Systems Analysis Simulation

The Cisco Networking Academy is a public–private partnership between Cisco and over 9,000 educational institutions in over 160 countries in which Cisco provides partnering schools with free online curriculum and online assessments to support local school instructors in teaching ICT skills in areas related to PC repair and maintenance, as well as computer and data network design, configuration, and maintenance in alignment with entry-level industry certifications.

A key skill in computer networking is troubleshooting. Cisco developed a simulation-based game using networking and entrepreneurial skills called Aspire. The main idea of Aspire is that students are entrepreneurs, starting their own small networking companies, and must make both business and technical decisions in the game. Aspire consists of a 2½-D interface that allows navigation, interaction with characters in the game, decision-making, and complex scenarios that combine numerous networking task requirements.

The simulation engine behind the game is called Packet Tracer (Frezzo, Behrens, & Mislevy, 2010). Packet Tracer is a domain-specific data network simulator used in Networking Academy curricula and performance-based assessments that provides instructional direction, practical experience, and assessment-based feedback throughout the courses. The Packet Tracer microworld supports a wide variety of networking devices, protocols, and interactions, giving the student ample opportunity for misconceptions, breakdowns, and suboptimal troubleshooting in addition to preferred approaches.

As described above, troubleshooting can be viewed as a type of systems-analysis problem. Jonassen and Hung (2006) described a number of subskills for troubleshooting, including:

- identifying fault symptoms;
- diagnosing faults;
- generating and verifying solutions.

In order to assess the ability to analyze complex systems in order to troubleshoot problems, in the Aspire game tasks were created with key features as follows:

- a configured network with a problem that makes it so information can not flow through it;
- methods available to identify the symptoms (i.e. inability of information to flow);
- methods available to diagnose the problem;
- multiple correct and incorrect solution paths;
- methods to test implemented solutions.

For example, one task placed the player in a medical office that had a network with four computers connected to a router. Players were told that three of the computers could not connect to the Internet while one could. Their task was to get them all connected. Students then had to use networking commands to identify which computer was working, investigate the network settings on the computers to determine which were faulty, and make the necessary corrections. There were two different ways to correct the incorrect settings in this particular problem.

The evidence model, as described in DiCerbo, Frezzo, and Deng (2011), broke down evidence for problem diagnosis, solution implementation, and solution evaluation:

SUBSKILLS OF TROUBLESHOOTING	EVIDENCE
Problem diagnosis	Use of Ping command prior to making any network changes (to determine which computer worked) Use of commands to display the network settings of computers
Solution implementation	Attempts to change the network settings of the computers or router
Solution evaluation	Use of Ping command after making network changes (to test whether computers worked)

Using this evidence, researchers were able to differentiate students who were novice versus expert troubleshooters (based on their enrollment in either the first or fourth semester of networking classes).

Evidence Model for an Argumentation Game

We can similarly build tasks and define evidence for argumentation-related skills. In the game Mars Generation One: Argubot Academy, players are students in the first colony on Mars. There are many debates about life on Mars, from what kind of protein to eat to whether there should be pets. These debates are solved through robot battles. Players must gather evidence then link claims and evidence to equip their robots for battle. In battle, their robots (i.e. claim-evidence pairings) are tested, and they must successfully identify weaknesses in their opponents' claim-evidence pairings. The tasks were to:

1. travel through the world gathering evidence,
2. create robots by linking evidence to relevant claims they wish to support;
3. engage in robot battles in which they critique others' arguments.

The evidence for their argumentation skill is shown below.

CRITICAL THINKING SUBSKILL	EVIDENCE
Argument analysis	Collection of evidence related to claims in environment Not collecting evidence unrelated to claims in the environment Correctly identify problems in opponents' claim–evidence pairings Number of opposing robots killed
Creation	Links multiple pieces of evidence to claims Uses multiple kinds of claims in battle Number of battles won

How Formative Feedback Can Be Used with ECD

When creating ECD models around critical thinking for the purpose of supporting learning, we need to consider the ways we give feedback about student performance as they complete tasks. There are two different types of feedback in the context of ECD:

1. student model feedback;
2. evidence model feedback.

Both are needed in a comprehensive feedback system, which is critical to supporting learning within a formative assessment process.

STUDENT MODEL FEEDBACK

Student model feedback is feedback around performance on a specific skill or subskill in a student model. For example, feedback may tell the student the percentage of activities they have already completed on a particular learning objective or subskill, or it may report some estimate of a student's level of mastery or proficiency on that subskill. This type of feedback is fairly shallow since it merely describes the performance on a skill rather than providing some indication of how the student could improve their performance. Although student model feedback may not support revision within an activity, research suggests that students find this type of feedback to be useful in understanding their progress over time (Loboda, Guerra, Hosseini, & Brusilovsky, 2014), that they regularly access this type of feedback when it is made available to them (Long & Aleven, 2011), and that students who are able to view their own student models show greater persistence and perform better than students who do not have access to their own student models (Falakmasir, Hsiao, Mazzola, Grant, & Brusilovsky, 2012; Loboda et al., 2014; Long & Aleven, 2013).

EVIDENCE MODEL FEEDBACK

Evidence model feedback is feedback around behaviors identified in the evidence model. This type of rich feedback can provide an indication of what behaviors are associated with improving skills in the student model – in Shute's (2008) language, it either tells the student exactly what aspects of their response need to be fixed (directive, which is especially good for novices) or suggests where there may be a problem and allows the student correct their own mistakes (facilitative, which may be better for more advanced learners). Feedback at the evidence level is more granular than feedback at the student model level, since specific behaviors can be targeted with recommendations for how to progress through a problem. Evidence model feedback is typically more difficult to implement at scale since it requires creating feedback for specific behaviors outlined in the evidence model.

Summary of Assessment of Critical Thinking

Traditional critical-thinking assessments are typically framed as domain-general. Domain-general critical-thinking assessments have been shown to predict important outcomes. The assessment of skills related to critical thinking in specific disciplines varies considerably and can be broadly grouped into writing, simulation, and concept-mapping tasks. In each task type there are multiple pieces of evidence that can be captured to inform the skills of critical thinking. The Jonassen problem types provide a way to create different problems to elicit critical-thinking skills.

Critical Thinking in Practice

Point/Counterpoint

One challenge in teaching educational psychology is helping students become critical thinkers about issues in educational-psychology research and practice. In Woolfolk's writing and teaching she has addressed this challenge through debates. Every chapter of her Educational Psychology text includes a Point/Counterpoint section that gives different sides of an argument. For example, in the chapter of children with disabilities, she includes the following:

Point/Counterpoint: Pills or Skills for Children with ADHD?

About 3% of school-age children in the United States (ages 6 to 18) take some kind of medication for ADHD. Should children with ADHD be given drugs?

POINT ▾

Yes, drugs are helpful in ADHD. About 30% of the people who take Ritalin, the most commonly prescribed ADHD drug, respond well (Hallahan et al., 2015). Ritalin and other prescribed drugs such as Adderall, Focalin, Dexadrine, Vyvanse, and Cylert are stimulants. In particular dosages that vary by individual, they appear to influence the release of neurotransmitters and help the executive functions of the brain to operate more normally (Hallahan et al., 2015). Short-term effects include possible improvements in behaviors, including increased cooperation, attention, task switching, and compliance. Research suggests that about 70% - 80% of children with ADHD are more manageable and better able to benefit from educational and social interventions when on medication (Hutchinson, 2009). In fact both stimulants such as Adderall and Ritalin and non-stimulant treatments such as Strattera appear to have some helpful effects for many children and adolescents with SDHD (Kratchovil, 2009). Positive results also have been reported with Buspar, typically used to treat anxiety, and even with some supplements such as pycnogenol (Trebaticka et al, 2009). And some evidence indicates that Strattera might have positive effects on working memory, planning, inhibition — at least for the Chinese children studied (Yang et al,, 2009).

COUNTERPOINT ▾

No, drugs should not be the first treatment tried with ADHD. Many children have negative side effects from drugs such as increased heart rate and higher blood pressure, interference with growth rate, insomnia, weight loss, and nausea (D.C. Smith et al., 2014). For most children, these side effects are mild and can be controlled by adjusting the dosage and timing of the drug. However, little is known about the long-term effects of drug therapy. The drug Strattera is not a stimulant, but use may lead to increased thoughts of suicide. As a parent or teacher, you need to keep up with the research on treatments for ADHD.

Many studies have concluded that the improvements in behavior resulting from the use of the drugs *seldom* lead to improvements in academic learning or peer relationships, two areas where children with ADHD have great problems. Because children appear to improve dramatically in their behavior, parents and teachers, relieved to see change, may assume the problem has been cured. It hasn't. The children still need special help in learning, especially interventions focused on task switching and how to make *connections* among elements in readings or presentations in order to build coherent, accurate representations of the information (Bailey et al., 2009; Doggett, 2004; Purdie, Hattie, & Carroll, 2002).

Beware of Either/Or. The bottom line is that even if students in your class are on medication, it is critical that they also learn the academic and social skills they will need to succeed — *this will not happen by itself*. Students need to learn how and when to apply learning strategies and study skills. Also, they need to be encouraged to persist when challenged by difficult tasks and to see themselves as having control over their learning and behavior. Medication alone will not make this happen, but it may help. We need to attack the problem on several fronts with effective teaching, counseling, and supports for positive behaviors.

Figure 1 Source: Woolfolk, forthcoming, p. 151.

Assessments

Please provide the pros and cons of using medication for the treatment of ADHD (Attention Deficit Hyperactivity Disorder). What would you choose if you had a child with ADHD?

Conclusions

CONCLUSIONS	RECOMMENDATION
Critical-thinking skills are highly sought after by employers and are associated with positive outcomes in many aspects of life.	Educators should include critical-thinking skills in their teaching.
Critical thinking involves the skills of argument analysis, systems analysis, creation, and evaluation.	Educators should target each of these aspects of critical-thinking in their instruction.
Teachers who have participated in extensive professional development in critical thinking have students who are better at critical thinking.	Schools and institutions should provide professional-development resources to teach critical-thinking in specific disciplines.
Models that explicitly teach critical thinking as part of teaching a specific discipline are effective.	Educators should situate explicit critical-thinking instruction in specific disciplines where applicable.
Problem-based learning, scaffolded practice, and collaborative learning approaches may be effective in promoting critical-thinking.	Educators should consider using problem-based learning, scaffolded practice, and collaborative learning approaches to teaching critical-thinking.
Activities such as concept maps, simulations, and structured argumentation practice have been shown to be effective forms of critical-thinking instruction, especially for systems and argument analysis.	Educators should consider using concept-mapping activities, simulations, and structured argumentation exercises to foster systems- and argument-analysis skills.
Problem types can be useful to help create critical-thinking assessments.	Educators should use the problem types as templates to help design or select appropriate critical-thinking assessments.
There are varying forms of feedback to foster learning in critical-thinking activities.	Instructors should provide feedback on student performance at both the skills level and at the evidence level to scaffold learning of critical-thinking skills

References

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.

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications*, 7(1), 39–59.

ACT. (no date). *Collegiate assessment of academic proficiency technical manual*. Iowa City, IA: ACT. Retrieved from www.act.org/content/dam/act/unsecured/documents/CAAP-TechnicalHandbook.pdf

Allen, J. A., Hayes, R. Y. T., & Buffardi, L. C. (2001). Maintenance training simulator fidelity and individual differences in transfer of training. In R. W. Sweezey and D. H. Andrews (Eds.), *Readings in training and simulation: A 30-year perspective* (pp. 272–284). Santa Monica, CA: Human Factors Society.

Alpert, S. R. (2006). Computer-based concept mapping. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 100–104). Hershey, PA: Idea Group.

American Institute of Certified Public Accountants. (2017) *Core competency framework.* Retrieved from http://www.aicpa.org/interestareas/accountingeducation/resources/pages/corecompetency.aspx

American Psychological Association. (2011). *Principles for quality undergraduate education in psychology.* Washington, DC: American Psychological Association. Retrieved from www.apa.org/education/undergrad/principles.aspx

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction*, 19(4), 322–334.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Association of American Colleges and Universities. (2011). *The LEAP vision for learning: Outcomes, practices, impact, and employers' view.* Washington, DC: Association of American Colleges and Universities.

Association of College and Research Libraries. (2015). Framework for information literacy for higher education. Retrieved from http://www.ala.org/acrl/standards/ilframework

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70, 181–214.

Bahr, N. (2010) Thinking critically about critical thinking in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), Article 9.

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

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 47–78). Boston, MA: Kluwer.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical-thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, 8(2), 25–42.

Behrens, J., DiCerbo, K., & Ferrara, S. (2012). Intended and unintended deceptions in the use of simulations. K-12 Center at ETS invitational research symposium on technology enhanced assessments. Retrieved from http://www.k12center.org/rsc/pdf/session2-pearson-behrens-dicerbo-ferrara-paper-tea2012.pdf

Bensley, D. A., & Haynes, C. (1995). The acquisition of general purpose strategic knowledge for argumentation. *Teaching of Psychology*, 22(1), 41–45.

Bensley, D. A., & Spero, R. A. (2014). Improving critical-thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity*, 12, 55–68.

Bernard, R., Zhang, D., Abrami, P., Sicoly, F., Borokhovski, E., & Surkes, M. (2008). Exploring the structure of the Watson–Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3(1), 15–22.

Blessing, S. B., & Blessing, J. S. (2010). PsychBusters: A means of fostering critical thinking in the introductory course. *Teaching of Psychology*, 37(3), 178–182.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company.

Braun, H., Bejar, I. I., & Williamson, D. M. (2006). Rule-based methods for automated scoring: Application in a licensing context. In D. M. Williamson, I. I. Bejar, & R. J. Mislevy (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 83–122). Mahwah, NJ: Lawrence Erlbaum Associates.

Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity*, 7(2), 112–121.

CAAP Program Management. (2012). *ACT CAAP technical handbook 2011–2012*. Iowa City, IA: CAAP Program Management. Retrieved from http://www.act.org/caap/pdf/CAAP-TechnicalHandbook.pdf

Cabrera, A. F., Colbeck, C. L., & Terenzini, P. T. (2001). Developing performance indicators for assessing classroom teaching practices and student learning. *Research in Higher Education*, 42(3), 327–352.

Calzada Prado, J., & Marzal, M. Á. (2013) Incorporating data literacy into information literacy programs: core competencies and contents. *Libri: International Journal of Libraries & Information Services*, 63(2): 123–134.

Carriger, M. S. (2016). What is the best way to develop new managers? Problem-based learning vs. lecture-based instruction. *International Journal of Management Education*, 14(2), 92–101.

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

Casner-Lotto, J., & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce. New York, NY: The Conference Board, Inc.

Catrambone, R., & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition*, 18(6), 593–603.

Cheng, B. H., Ructtinger, L., Fujii, R., & Mislevy, R. (2010). *Assessing systems thinking and complexity in science (large-scale assessment technical report 7)*. Menlo Park, CA: SRI International.

Chi, M. T. H., & Van Lehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences*, 1(1): 69–105.

Cho, K. L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology: Research & Development*, 50(3), 5–22.

Consorti, F., Mancuso, R., Nocioni, M., & Piccolo, A. (2012). Efficacy of virtual patients in medical education: A meta-analysis of randomized studies. *Computers & Education*, 59, 1001–1008.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201.

de Klerk, S., Veldkamp, B., & Eggen, T. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, 85, 23–34.

DiCerbo, K. E., & Behrens, J. T. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 273–306). Charlotte, NC: Information Age Publishing.

DiCerbo, K. E., Frezzo, D. C., & Deng, T. (2011). Substantive validity of a simulation-based game. *Research and Practice in Technology Enhanced Learning*, 6(3), 161–185.

Educational Testing Service. (2010). *ETS proficiency profile user's guide*. Princeton, NJ: Educational Testing Service.

—— (2013). *Quantitative market research* [PowerPoint slides]. Princeton, NJ: Educational Testing Service.

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

—— (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10.

Ennis, R. H., Millman, J., & Tomko, T. N. (1985). *Cornell critical thinking tests*. Pacific Grove, CA: Midwest Publications.

Ennis, R. H., & Weir, E. (1985). *The Ennis–Weir critical thinking essay test.* Pacific Grove, CA: Midwest Publications.

Erdogan, Y. (2009). Paper-based and computer-based concept mappings: The effects on computer achievement, computer anxiety and computer attitude. *British Journal of Educational Technology*, 40(5), 821–836.

Ericsson, A. (2016). *Peak: Secrets from the new science of expertise.* Boston, MA: Houghton Mifflin Harcourt.

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

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

Falakmasir, M. H., Hsiao, I. H., Mazzola, L., Grant, N., & Brusilovsky, P. (2012). The impact of social performance visualization on students. *Proceedings of the IEEE 12th International Conference on Advanced Learning Technologies*, July 4–6, 565–569.

Foltz, P. W. (2015). Advances in automated scoring of writing for performance assessment. In Y. Rosen, S. Ferrara, & M. Mosharref (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 659–678). Hershey, PA: IGI Global.

Foltz, P. W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in online writing evaluation with LSA. *Interactive Learning Environments*, 8(2), 111–129.

Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2010). Design patterns for learning and assessment: Facilitating the introduction of a complex simulation-based learning environment into a community of instructors. *Journal of Science Education and Technology*, 19(2), 105–114.

Funke, J. (2003). Problemlösendes Denken [*Problem solving thinking*]. Stuttgart: Kohlhammer.

Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research & Development*, 52(2), 5–22.

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology*, 45(6), 1097–1114.

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991–2000. *Journal of College Student Development*, 44(6), 746–762.

Germann, P. J. (1989), Directed-inquiry approach to learning science process skills: treatment effects and aptitude-treatment interactions. *Journal of Research in Science Teaching*, 26(3): 237–250.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476.

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38–43.

Guyton, E. M. (1988). Critical thinking and political participation: Development and assessment of a causal model. *Theory & Research in Social Education*, 16(1), 23–49.

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

—— (2001) Assessing the effectiveness of critical thinking instruction. *Journal of General Education*, 50(4), 270–286.

—— (2003). Thought and knowledge: *An introduction to critical thinking*. Mahwah, NJ: Erlbaum.

—— (2010). Halpern *Critical Thinking Assessment manual*. Vienna: Schuhfried GmbH.

Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction*, 29, 31–42.

Hernandez-Serrano, J., & Jonassen, D. H. (2003). The effects of case libraries on problem solving. *Journal of Computer Assisted Learning*, 19, 103–114.

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, 86(2), 431–468.

Hudgins, B. B., & Edelman, S. (1988). Children's self-directed critical thinking: A model for its analysis and two examples. *Journal of Educational Research*, 81(5), 262–273.

Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In P. Dillenbourg (Ed.), *Learning to argue*, Vol. I (pp. 205–226). Dordrecht: Kluwer.

Johnson, W. B., & Norton, J. E. (1992). Modeling student performance in diagnostic tasks: A decade of evolution. In J. W. Regian and V. J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 195–216), Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson, W. B., & Rouse, W. B. (2001). Training maintenance technicians for troubleshooting: Two experiments with computer simulations. In R. W. Sweezy and D. H. Andrews (Eds.), *Readings in training and simulation: A 30-year perspective*, Santa Monica, CA: Human Factors Society.

Jonassen, D. H. (1997). Instructional design model for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development*, 45(1), 65–95.

—— (2000). Toward a design theory of problem solving. *Educational Technology: Research & Development*, 48(4), 63–85.

—— (2004). *Learning to solve problems: An instructional design guide*. San Francisco, CA: Wiley.

—— (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. London and New York: Routledge.

Jonassen, D. H., & Hung, W. (2006). Learning to troubleshoot: A new theory-based design architecture. *Educational Psychology Review*, 18, 77–114.

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology: Research and Development*, 58(4), 439–457.

Jonassen, D. H., & Hernandez-Serrano, J. (2002). Case-based reasoning and instructional design: Using stories to support problem solving. *Educational Technology: Research and Development*, 50(2), 65–77.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus & Giroux.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.

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, NJ: Lawrence Erlbaum Associates.

Klein P., Olson D. R., & Stanovich K. (1997). Structuring reflection: Teaching argument concepts and strategies enhances critical thinking. *Canadian Journal of School Psychology*, 13(1), 38–47.

Kobrin, J., Sato, E., Lai, E., & Weegar, J. (2016). Examining the constructs assessed by published tests of critical thinking. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.

Koenig, A. D., Lee, J. J., Iseli, M. R., & Wainess, R. (2010). A conceptual framework for assessing performance in games and simulation (CRESST Report 771). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Kollar, I., Fischer, F., & Slotta, D. J. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, 17(6), 708–721.

Kong, L. N., Qin, B., Zhou, Y. Q., Mou, S. Y., & Gao, H. M. (2014). The effectiveness of problem-based learning on development of nursing students' critical thinking: A systematic review and meta-analysis. *International Journal of Nursing Studies*, 51(3), 458–469.

Kong, S. C. (2014). Developing information literacy and critical-thinking skills through domain knowledge learning in digital classrooms: An experience of practicing flipped classroom strategy. *Computers & Education*, 78, 160–173.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16–25.

Kuhn, D., Hemberger, L., & Khait, V. (2015). *Argue with me: Argument as a path to developing students' thinking and writing*, 2nd edn. London and New York: Routledge.

Kuncel, N. R. (2011). Measurement and meaning of critical thinking. Report presented at the National Research Council's 21st Century Skills Workshop, Irvine, CA, January.

Lee, H.-S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education*, 24(2), 115–136.

Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with the California Critical Thinking Tests. Nurse Education, 22(5), 29–33.

Levy, R. (2013). Psychometric and evidentiary advances, opportunities, and challenges for simulation-based assessment. *Educational Assessment*, 18(3), 182–207.

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

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

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next-generation assessment* (ETS RR-14–10). Princeton, NJ: Educational Testing Service.

Loboda, T. D., Guerra, J., Hosseini, R., & Brusilovsky, P. (2014). Mastery grids: An open source social educational progress visualization. In *European Conference on Technology Enhanced Learning*, 235–248.

Loes, C. N., & Pascarella, E. T. (2017). Collaborative learning and critical thinking: Testing the link. *Journal of Higher Education*, 88(1): 1–28.

Loes, C., Pascarella, E., & Umbach, P. (2012). Effects of diversity experiences on critical-thinking skills: Who benefits? *Journal of Higher Education*, 83(1), 1–25.

Long, Y., & Aleven, V. (2011). Students' understanding of their student model. *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, June, 179–186.

Lovelace, K. J., Eggers, F., & Dyck, L. R. (2016). I do and I understand: Assessing the utility of web-based management simulations to develop critical-thinking skills. *Academy of Management Learning & Education*, 15(1), 100–121.

Macionis, J. J. (2003). Sociology (9th ed.). Upper Saddle River, NJ: Prentice-Hall.

Margolis, M. J., & Clauser, B. E. (2006). A regression-based procedure for automated scoring of a complex medical performance assessment. In D. M. Williamson, I. I. Bejar, & R. J. Mislevy (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 123–168). Mahwah, NJ: Lawrence Erlbaum Associates.

Marin, L., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6, 1–13.

Markle, R., Brenneman, M., Jackson, T., Burrus, J., & Robbins, S. (2013). *Synthesizing frameworks of higher education student learning outcomes*. Princeton, NJ: Educational Testing Service.

Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom: A longitudinal case study. *Research in Science Education*, 39(1), 17–38.

Mayer, B. W., Dale, K. M., Fraccastoro, K. A., & Moss, G. (2011). Improving transfer of learning: relationship to methods of using business simulation. *Simulations and Gaming*, 42(1), 64–84.

McAllister, S., Ravenscroft, A., & Scanlon, E. (2004). Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC. *Journal of Computer Assisted Learning*, 20(3), 194–204.

McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475–492.

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

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–67.

Mislevy, R. J., Steinberg, L. S., Almond, R. G., Breyer, F. J., & Johnson, L. (2001). *Making sense of data from complex assessment* (CRESST Report 538). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical-thinking skills? A meta-analysis. *Educational Research Review*, 9, 114–128.

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL): A systematic review and synthesis of fifteen years of research. *Educational Research Review*, 7(2), 79–106.

Nussbaum, E. M., Hartley, K., Sinatra, G. M., Reynolds, R. E., & Bendixen, L. D. (2004). Personality interactions and scaffolding in on-line discussions. *Journal of Educational Computing Research,* 30(1 & 2), 113–137.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, 97(2), 157–169.

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, 76(1), 59–92.

OECD. (2012). *Education at a Glance 2012: OECD Indicators*. Paris: OECD Publishing.

O'Keefe, D. J. (1998). Justification explicitness and persuasive effect: A meta-analytic review of the effects of varying support articulation in persuasive messages. *Argumentation and Advocacy*, 35, 61–75.

Park, O. K., & Gittelman, S. S. (1992). Selective use of animation and feedback in computer-based instruction. *Educational Technology Research. Development, 40(4)*: 27–38.

Partnership for 21st Century Learning. (P21) (2015). Framework for 21st Century Learning. Retrieved from http://www.p21.org/our-work/p21-framework

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students, vol. II: A third decade of research.* San Francisco, CA: Jossey-Bass.

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

Penningroth, S. L., Despain, L. H., & Gray, M. J. (2007). A course designed to improve psychological critical thinking. *Teaching of Psychology*, 34(3), 153–157.

Quitadamo, I. J., Brahler, C. J., & Crouch, G. J. (2009). Peer-led team learning: A prospective method for increasing critical thinking in undergraduate science courses. *Science Educator*, 18(1), 29–39.

Ravenscroft, A. (2011). Dialogue and connectivism: A new approach to understanding and promoting dialogue-rich networked learning. *International Review of Research in Open and Distance Learning*, 12(3), 139–160.

Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In L. Gleitman & A. K. Joshi (Eds.), *Proceeding of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393–398). Mahwah, NJ: Lawrence Erlbaum Associates.

Riesenmy, M. R., Mitchell, S., Hudgins, B. B., & Ebel, D. (1991). Retention and transfer of children's self-directed critical thinking skills. *Journal of Educational Research*, 85, 14–25.

Rimiene, V. (2002). Assessing and developing students' critical thinking. *Psychology Learning & Teaching*, 2(1), 17–22.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93(2), 346–362.

Rowe, M. P., Gillespie, B. M., Harris, K. R., Koether, S. D., Shannon, L.-J. Y., & Rose, L. A. (2015). Redesigning a general education science course to promote critical thinking. *CBE-Life Sciences Education*, 14(3), 1–12.

Ruiz-Primo, M. A. (2000). On the use of concept maps as an assessment tool in science: What we have learned so far. *Revista Electronica de Investigacion Educativa*, 2(1), 29–53.

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate from being scripted or from observing a model. *International Journal of Computer-Supported Collaborative Learning*, 26(4), 69–92.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment*, 8(4), 4–47.

Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered design for diagnostic assessment within digital learning environments: Integrating modern psychometrics and educational data mining. *Journal of Educational Data Mining*, 4(1), 1–10.

Schaal, S., Bogner, F. X., & Girwidz, R. (2010). Concept mapping assessment of media assisted learning in interdisciplinary science education. *Research in Science Education*, 40(3), 339–352.

Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchoronous discussion groups? *International Journal of Computer-Supported Collaborative Learning*, 2(2–3), 225–246.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artifact of lousy control conditions. *Computers in Human Behavior*, 25(2), 258–266.

Şendağ, S., & Odabaşı, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical-thinking skills. *Computers & Education*, 53(1), 132–141.

Shemwell, J., Fu, U., Figueroa, M., Davis, R., & Shavelson, R. (2010). Assessment in schools: Secondary science. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education*, 3rd edn (vol. III, pp. 300–310). Amsterdam: Academic Press.

Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and future directions*. London and New York: Routledge.

Shermis, M. D., & Hamner. B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis. Paper presented at Annual Meeting of the National Council on Measurement in Education, Vancouver, Canada, April.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.

Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: The MIT Press.

Siewiorek, A., Gegenfurtner, A., Lainema, T., Saarinen, E., & Lehtinen, E. (2013). The effects of computer simulation game training on participants' opinions on leadership styles. *British Journal of Educational Technology*, 44(6), 1012–1035.

Song, Y., Deane, P., Graf, E. A., & van Rijn, P. (2013). *Using argumentation learning progressions to support teaching and assessments of English language arts.* Princeton, NJ: Educational Testing Service.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive processing in computer-supported collaborative learning environment. *Instructional Science*, 40(2), 297–323.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2(4), 421–447.

Steiner, C. M., Albert, D., & Heller, J. (2007). Concept mapping as a means to build e-learning. In N. A. Buzzetto-More (Ed.), *Advanced principles of effective e-learning* (pp. 59–111). Santa Rosa, CA: Informing Science Press.

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

Stoyanov, S., & Kommers, P. (2008). Concept mapping instrumental support for problem solving. *International Journal Continuing Engineering Education and Life-Long Learning,* 18(1), 40–53.

Sweeney, L. B., and Sterman, J. D. (2007). Thinking about systems: Student and teacher conceptions of natural and social systems. *System Dynamics Review*, 23 (2–3), 285–312.

Thayer-Bacon, B. J. (2000). *Transforming critical thinking: Thinking constructively.* New York, NY: Teachers College Press.

Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 53(1), 41–48.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, 36(3), 212–218.

Ventura, M., Shute, V., Wright, T., & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology*, 4, 852.

Walker, A., & Leary, H. (2009). A problem based learning meta-analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, 3(1), 6–28.

Watson, G., & Glaser, E. M. (1980). *Watson–Glaser Critical Thinking Appraisal, forms A and B manual.* San Antonio, TX: The Psychological Corporation.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning & Instruction*, 17(4), 416–426.

Weinerth, K., Koenig, V., Brunner, M., and Martin, R. (2014) Concept maps: A useful and usable tool for computer-based knowledge assessment? A literature review with a focus on usability. *Computers & Education*, 78, 201–209.

West, P., Wise Rutstein, D., Mislevy, R. J., Junhui, L., Levy, R., Dicerbo, K. E., Crawford, A. . . . & Behrens, J. T. (2010). *A Bayesian network approach to modeling learning progressions and task performance (CRESST Report 776).* Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Willingham, D. T. (2007). Critical thinking. *American Educator,* 31(3), 8–19.

Wing, J. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–36.

—— (2011). Research notebook: Computational thinking—What and why? *The Link Newsletter,* 6, 1–32.

Woolfolk, A. (forthcoming). *Educational psychology*, 14th edn. Boston, MA: Pearson.

Yiong-Hwee, T., & Churchill, D. (2007). Using sentence openers to support students' argumentation in an online learning environment. *Educational Media International*, 44(3), 207–218.

Yue, M., Zhang, M., Zhang, C., & Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 52, 87–94.

Zabit, M. N. M., Karagiannidou, E., & Zachariah, T. Z. (2016). Teaching business in Malaysia and the use of PBL to nurture students' critical thinking: A case study of Sultan Idris Education University. *Malaysian Journal of Society and Space*, 12(6): 1–11.

Zahner, D. (2013). *Reliability and validity: CLA+*. New York, NY: Council for Aid to Education.

Zohar, A., & Tamir, P. (1993). Incorporating critical thinking into a regular high school biology curriculum. *School Science and Mathematics*, 93: 136–140.

ALWAYS LEARNING