

CAE

The Case for Performance-Based Assessment of Critical-Thinking Skills Multiple-Choice Tests Shift From a Dominant to a Supporting Role

November 6, 2014



Introduction

All approaches to education assessment have costs and benefits. Take, for example, the multiple-choice paradigm that has dominated educational assessment in the K-12 and post-secondary education systems. Supporters of multiple-choice assessment have long argued that selected-response items are efficient to administer, easy to grade, and inexpensive for participating institutions. Moreover, they do a credible job of assessing the constructs they are intended to measure. In today's Knowledge Economy, however, there is increasing agreement that critical-thinking, problem-solving, qualitative and quantitative analytic-based reasoning, and written-communication skills need to be improved. The multiple-choice test may do a satisfactory job of measuring student mastery of content; however, mastery of content is not enough to survive and prosper in the 21st century. In fact, the amount of available information creates a higher-order problem.

Individuals need to be able to access, structure, and use all of the available information. This places a premium on critical-thinking skills. Here, multiple-choice tests are not adequate. Constructed responses are superior because they simulate the higher-order skills needed for success. The performance-based approach to assessment requires students to demonstrate proficiency in the critical-thinking skills they have achieved as a function of taking the test. Educators and assessment experts have known about performance assessments but have considered them too costly, error prone, and not reliable enough to be used widely. Recent advancements in education technology, stimulated by Internet-based solutions, have changed this equation dramatically. Performance-based assessments are now poised to be the dominant paradigm in educational assessment throughout the K-12 and post-secondary education systems. While selected-response items can be used to support performance-based assessments (i.e., to ensure reliability for high-stakes decisions about individual students), their dominant position is a thing of the past. This paper elaborates on the need for performance-based assessment.

Background

Educational institutions across the world are being challenged to improve instruction so that tomorrow's workforce will have the knowledge and skills necessary to meet the demands of modern careers while contributing to the global economy. Indeed, a college education has never been more necessary for productive participation in society. Employers now seek individuals who are able to think critically and communicate effectively in order to meet the requirements of the new Knowledge Economy. Therefore, the skills taught in higher education are changing; less emphasis is placed on content-specific knowledge and more emphasis is placed on critical-thinking skills, such as analytic and quantitative reasoning, problem-solving, and written communication.

Any rigorous improvement project requires constant evaluation in order to measure progress toward goals. Consequently, there is a clear need for standardized assessments that measure improvement in student learning (i.e., critical-thinking skills). Performance-based assessments, such as CLA+, not only evaluate whether students are learning the critical-thinking skills required of today's workforce, they also spur advances in educational pedagogy. The CLA+ protocol is a 90-minute test that is anchored by a performance task designed to take 60 minutes and is accompanied by 25 selected-response items; these items are similar in form to multiple-choice tests but are designed to be scaffolded, requiring active participation by the test-taker to examine statements and connect the dots between arguments in order to select the appropriate response. Moreover, the selected-response questions are aligned with the content of the performance assessment.

Two competitors to CLA+ claim to measure critical-thinking skills solely with multiple-choice questions (Proficiency Profile by Educational Testing Service (ETS) and the Collegiate Assessment of Academic Proficiency by ACT). CLA+ Performance Tasks, however, present

students with scenarios that are representative of the types of problems they will encounter in the real world and require them to generate their own solutions to these problems. Traditional multiple-choice questions have limited capacity to measure students' critical-thinking skills. However, open-ended assessments, such as CLA+, are able to measure how well students formulate hypotheses, recognize fallacious reasoning, and identify implicit and possibly incorrect assumptions.

The study of the return on the amount invested in a student's education has become an important pursuit in economic and education policy. This body of research suggests that education must focus on the stock of knowledge and skills required by a society. Today's society highly values a person's ability to access and structure information and then apply that information to solve new problems—critical thinking. Increasing recognition of the essential role of these critical-thinking skills in the Knowledge Economy portends significant changes in teaching and learning. This is currently reflected in the deeper-learning reform movement underway in K-12 and post-secondary education. The reform movement in education in the United States can be characterized along the following three goals:

1. Shifting from the long-standing lecture format to a student-centered approach, emphasizing students' active class participation and development of analytic writing skills
2. Changing the balance of curriculum and textbook focus, from their current emphasis on memorizing content, to a focus on case studies and problem-based materials that require students to apply content to novel situations
3. Changing assessment instruments from multiple-choice tests that are best left for benchmarking the level of content absorbed by students to open-ended assessments that measure the whole student and are aligned with the goals of the reform initiative

Although significant advances have been made on the first two characterizations of this education reform, assessment of a student's critical-thinking skills has lagged behind. As schools and colleges focus increasingly on developing critical-thinking skills in their students, assessment tools will need to evolve to properly measure how well students are learning—and how effectively the institutions are teaching—such skills.

The Role of Performance-Based Assessment for Measuring Critical Thinking in the Knowledge Economy

Today's Knowledge Economy presents significantly different opportunities and challenges than those posed by the Industrial Era and creates a need for an education reform movement in K-12 and post-secondary education. This new environment places an emphasis on change, the production of services, and the creation of intellectual content and knowledge. This puts institutions, which rely on continuity and tradition, in a challenging position.

The quality of service, whether in health care, education, or even an app, becomes more important than the service's simple existence. Moreover, the pace of economic and social change, including the decline or restructuring of the largely centralized institutions from the past, creates the need to be able to apply what one knows to new situations. In other words, it becomes essential to be able to access, structure, and use information which, in and of itself, has become a primary issue to deal with because of "information overload." Therefore, while acquiring content remains important, a new requirement is the need to improve critical thinking, problem-solving, analytic and quantitative reasoning, and written communication so one can apply the vast quantities of information available to the problem at hand. In K-12 education, this goal is expressed by the call for deeper learning by the Common Core movement. With respect to post-secondary education, the three goals are expressed by the reform movement mentioned above.

Performance-based assessment is the best solution for measuring higher-order skills. Critical thinking is an integrative activity that requires active, not passive, participation by those looking to succeed. The premise of the performance-based assessment paradigm is that the whole is greater than the sum of the parts. An easy example is demonstrated by one seeking a driver's license. There, the applicant not only takes a short multiple-choice test but must also demonstrate an ability to drive by actually driving, all while being observed by an examiner.

The examiner may ask the applicant to, for example, turn left in two blocks and then parallel park. To carry out these steps the applicant must make a significant number of decisions in an integrated manner. Likewise, in the Knowledge Economy, many, if not most, daily needs require performing an activity in order to solve a problem and not just passively selecting the “right” response.

Multiple-choice tests remain the dominant testing regime, not only for facts, but also—albeit ineffectively—for critical-thinking skills. In the United States, multiple-choice tests are used overwhelmingly by ETS, ACT, and the College Board. As a result, the post-secondary education testing regime is not only failing to properly assess the most crucial skills required of students in the workplace but, just as importantly, is also failing to support the other two goals of the reform movement. CAE believes that the promise of educational reform cannot be achieved without employing open-ended, performance-based assessments in post-secondary education, as well as in primary and secondary education.

As an illustration of this point, consider two tests of critical thinking: one multiple-choice and the other a performance-based assessment—both aimed at measuring students’ understanding of correlations and causality.

On the one hand, a multiple-choice test, when used on its own, provides a list of four or five options and asks the student to select an answer. A performance-based assessment, on the other hand, presents students with a research report where, for example, the author incorrectly concluded that there is a causal relationship between two variables due to a seemingly strong correlation between them. But here, the student is not asked to merely select from options on a menu; rather, the student must first evaluate the information and then determine whether and how that information supports possible solutions to a realistic problem.

The cognitive processes involved in responding to the two contrasting assessments are fundamentally different. Recognizing the correct answer from a finite list of possibilities differs greatly from asking students to generate a critique and explain it. In the latter approach, the student must not only recognize the fallacious reasoning but must also understand how the concepts are confused and explain why the argument fails. This level of fidelity to real-world experience is often viewed as a major advantage of performance-based assessments over multiple-choice tests. Additionally, performance-based assessments measure a student’s written-communication skills and ability to use relevant and reliable information in crafting arguments and refuting counter arguments. Multiple-choice items that assess writing generally measure no more than a student’s ability to simply identify proper use of vocabulary and grammar.

Another important advantage of performance-based assessments is that they are regarded as tests worth teaching to. The practice of “teaching to the test” is generally frowned upon when referring to traditional multiple-choice and short-answer assessments. Unfortunately, there is ample evidence that this practice not only occurs but is, by its very nature, encouraged. When educators are held accountable for their students’ test performance, how else can a teacher be expected to succeed? However, “teaching to the test” should be encouraged for performance-based assessments. Class time spent teaching students how to apply knowledge and skills to complex, real-world problems is time well spent. Integrating performance-based assessments into accountability systems has the potential to positively impact classroom practice by encouraging teachers to foster the development of the competencies defined as critical-thinking skills.

In addition to negative effects on pedagogy, another critical shortcoming of today’s multiple-choice assessment regime is that it pays little attention to how much a school or college contributes to developing the competencies students will need after graduation. For instance, the criteria that are typically looked at by higher-education accreditation teams—a college’s retention rate, graduation rate, and percentage of tenured faculty—say nothing about how well the school fosters the development of its students’ analytic reasoning, problem-solving, and written-communication skills. This situation is unfortunate because, as demonstrated above, with respect to teachers, institutional evaluation and ranking significantly affect the priorities of an institution. If institutions were held accountable for achievements in student learning, they would likely direct greater resources and effort toward improving a student’s ability to think critically.

CLA+ Performance-Based Assessments Versus Multiple-Choice Tests

CLA+ assessments are comprised of performance tasks and enhanced selected-response questions. The CLA + Performance Tasks (PTs) are presented in a variety of contexts, including the arts, social sciences, natural sciences, business, education, and political science, among many other fields. That said, no prior subject-matter knowledge is required. Rather, students use their analytical reasoning, problem-solving, and written-communication skills to answer open-ended questions that are not framed to elicit “right” or “wrong” answers. Students are instead asked to compose written responses wherein they integrate information from provided reference documents and draw support for their decisions with relevant facts and ideas from those documents. The enhanced selected-response questions are aligned with a PT. Taken together, PTs and selected-response questions provide the most authentic and actionable measure of critical-thinking outcomes. This structure increases the reliability of individual student results.

While the differences between the CLA+ performance-based assessment and multiple-choice tests are many, the shortcomings of multiple-choice tests can be summarized as follows:

- The individual multiple-choice items only take on the appearance of the construct being measured.
- Multiple-choice tests are faster, cheaper, and easier to grade but cannot replicate the quality and authenticity of a performance task.
- Performance tasks are as reliable as multiple-choice tests and imitate demands and challenges that students will encounter in their post-secondary lives.

These points are elaborated on below.

The fallacy of multiple-choice tests is that a suite of individual questions (items) must be assumed to add up to a whole (i.e., the construct being measured). This is an irrational conclusion given the demands of the Knowledge Economy. Perhaps this made sense in the Industrial Era when the production of physical goods was the dominant output for the economy, and the growth of information was in proportion with the relatively slow growth of the largely centralized economic, social, and political institutions of the time. Moreover, it may have been more reasonable to assume continuity in those institutions and their rate of growth. As such, a society focused primarily on the production of physical goods translated into an educational need for assessments that probed the ability of the student to remember facts. The popularity of multiple-choice assessments in the 20th century reflects the economic and societal demands of the time. This also meant that it was reasonable to simply ask test-takers to correctly identify the “right” answer to relevant factual questions (e.g., What do steamships carry? (a) cargo; (b) people; (c) both; (d) none of the above).

While recognizing the importance of content knowledge, CAE asserts that higher-order skills are necessary for success in the 21st century and that performance tasks are the most appropriate instrument to measure these skills.

The CLA+ protocol is a 90-minute test that is anchored by a performance task designed to take 60-minutes and is accompanied by 25 enhanced selected-response questions that increase the reliability of the protocol to a level comparable to multiple-choice tests. A performance-based task and set of selected-response questions, which, when taken together, offer an authentic, actionable, and reliable measure of student-learning outcomes. In fact, the enhanced selected-response questions are innovative versions of the multiple-choice form, requiring active participation on the part of the student before selecting the “right” answer. Additionally, these selected-response questions align with the performance task.

Although multiple-choice items can be developed to measure the same construct (i.e., critical-thinking skills) as open-ended, performance-based assessments, there are differing levels of cognitive complexity. For example, when measuring critical reading and evaluation, multiple-choice items can assess whether students understand the main purpose of a passage. While this is a “document” based activity and the skill is important and necessary, it assesses a general understanding of a given text or passage, which CAE considers superficial.

Critical-thinking items can also measure logic and deductive reasoning, which are more abstract in nature but represent a very specific type of reasoning. Students certainly need to reason through the question before answering it but that does not necessarily reflect the type of reasoning CLA+ assesses. Students need to solve a logic problem, and logic—although not content specific—does not encompass the type of reasoning skills that can be generalized over numerous multiple-choice items.

CLA+ selected-response questions (SRQs), which are aligned to the same constructs as the PTs, require students to evaluate pertinent information in the form of informational or literary texts, graphs, tables, and/or diagrams. SRQs are more cognitively complex tasks since students are required to do what they would for the PTs: read the provided documents and answer questions that require them to analyze, integrate, and evaluate information to solve a problem. This is also more representative of the type of reasoning skills that are generalizable to multiple contexts, such as a workplace. Like the CLA+ PT, the CLA+ SRQ section, is representative of an authentic real-world type of task. **The combination of CLA+ SRQs and PTs is the most authentic, complete, and reliable way of measuring critical-thinking skills.**

The final defense of advocates of the multiple-choice paradigm is the assertion that performance-based assessments are much more expensive than multiple-choice tests. Multiple-choice tests are faster, cheaper, and easier to grade. When performance-based assessments were initially introduced into post-secondary education, they were, admittedly, expensive to develop, administer, and score. However, advances in item development, test administration, and scoring, including computer-assisted scoring, has lowered the cost of performance-based assessments significantly—to within the range of the cost for their multiple-choice counterpart.

Most significantly, CAE challenges the claim that “faster and cheaper” equates to a better assessment or even an assessment that meets a minimum requirement of involvement from the student. By promoting a simplified 30-minute set of items on “critical thinking,” one will never have a true measure of a student’s critical-thinking skills.

If multiple-choice items purport to capture the critical-thinking cognitive domain, one can be assured that the sum of the parts will not add up to a greater whole. Rather, one should use an assessment that captures the integrative nature of critical thinking itself. This means a protocol that is anchored by an open-ended constructed response is an essential component of the assessment. The student must be presented with the task of applying information and creating an appropriate response to a realistic problem. This is the only way true “critical thinking” can be adequately demonstrated.