

Criteria for High-Quality Assessment

June 2013

Linda Darling-Hammond, Joan Herman, James Pellegrino, Jamal Abedi, J. Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, Kenji Hakuta, Andrew Ho, Robert Lee Linn, P. David Pearson, James Popham, Lauren Resnick, Alan H. Schoenfeld, Richard Shavelson, Lorrie A. Shepard, Lee Shulman, Claude M. Steele

Soon, many states and districts will develop or select new assessments aligned with new student-learning standards they have adopted. These assessments may include those created by multi-state consortia as well as other state or local instruments. This brief is intended to support assessment developers and policymakers as they work to create and adopt assessments that promote deeper learning of 21st century skills.

The United States is poised to make a major shift in how we educate our children, increasing the focus of instruction on higher-order “21st century” skills that can help students succeed in a knowledge-based society. Today, college and career readiness depend much less on memorizing information, and much more on young people’s ability to analyze and apply what they’ve learned to address new problems, design solutions, collaborate effectively, and communicate persuasively.

There is no doubt that skill demands have changed. In 1970, for example, Fortune 500 companies were calling for the 3 Rs: reading, writing, and arithmetic. By the turn of the century, however, these companies were prioritizing teamwork, problem solving, and interpersonal skills. (See Figure 1.)

One catalyst for the shift in educational focus is the adoption of the Common Core State Standards (CCSS) by more than 40 states. The new standards — intended to be “fewer, higher, and deeper” than many states’ earlier standards — are designed to be internationally comparable and aligned with the changing nature of work and society.

As states implement these standards, they must also develop effective ways to evaluate how well the standards are being taught. Fill-in-the-bubble tests will not do the trick. As other nations have already discovered, new performance assessments are needed to evaluate how well students can find, evaluate, and use information rather than just recall facts. In high-achieving countries, these assessments frequently call on students to demonstrate what they know through written, oral, mathematical, physical, and multimedia products.

As explained in a recently-released report by the Gordon Commission, representing many of the nation’s leading experts in curriculum, teaching, and assessment:

New assessments must advance competencies that are matched to the era in which we live. Contemporary students must be able to evaluate the validity and relevance of disparate pieces of information and draw conclusions from them. They need to use what they know to make conjectures and seek evidence to test them, come up with new ideas, and contribute productively to their networks, whether on the job or in their communities. As the world grows increasingly complex and interconnected... the emphasis in our educational systems needs to be on helping individuals make sense out of the world and how to operate effectively within it.

Figure 1: Fortune 500 Most Valued Skills

	1970	1999
1	Writing	Teamwork
2	Computational Skills	Problem Solving
3	Reading Skills	Interpersonal Skills
4	Oral Communications	Oral Communications
5	Listening Skills	Listening Skills
6	Personal Career Development	Personal Career Development
7	Creative Thinking	Creative Thinking
8	Leadership	Leadership
9	Goal Setting / Motivation	Goal Setting / Motivation
10	Teamwork	Writing

What Should High-Quality Assessment Systems Include?

No single assessment can evaluate all of the kinds of learning we value for students or meet all of the goals held by parents, practitioners, and policymakers. In a coordinated system of assessments, different tools should be used for different purposes: formative and summative, diagnostic, and large-scale reporting. However, all assessments should faithfully represent the Standards and model good teaching and learning practice. We urge that systems be evaluated by these five criteria:

1: Assessment of Higher-Order Cognitive Skills

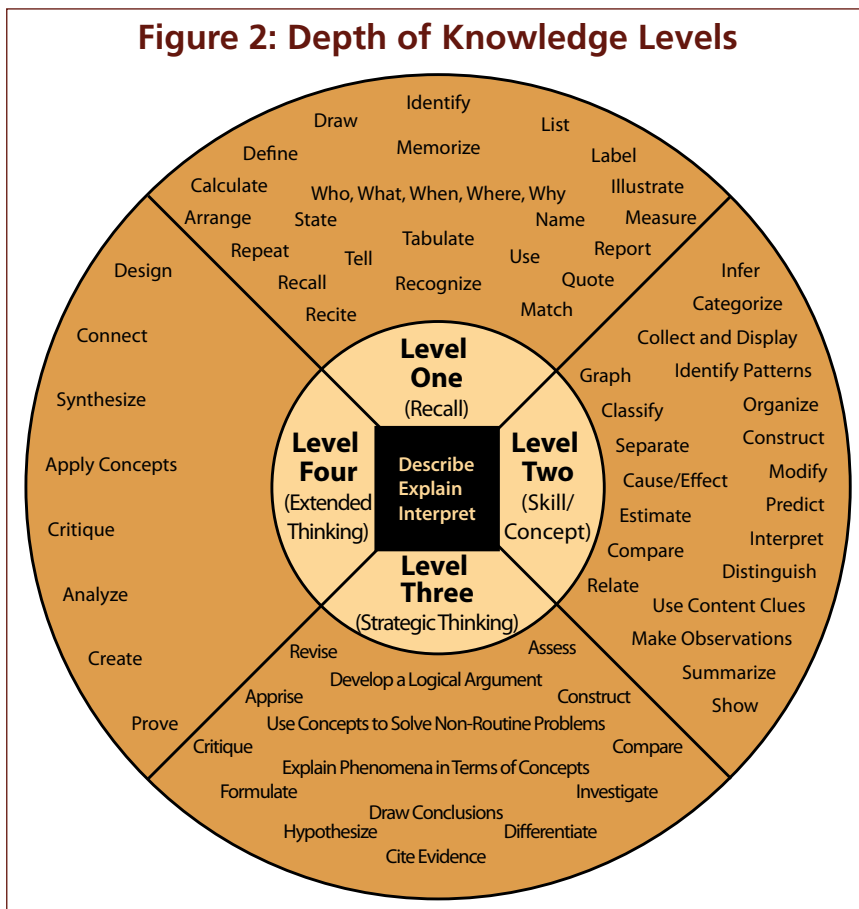
New assessments should tap the “higher-level” cognitive skills that allow students to transfer their learning to new situations and problems. These skills are rarely measured by current U.S. tests. A recent study of tests in 17 states found that fewer than 2% of mathematics items and only 21% of English language arts items addressed higher-level skills, such as those represented at the upper levels of the Depth of Knowledge (DOK) framework shown in Figure 2, below. These include, for example, the abilities to evaluate, compare, hypothesize, and investigate (Level Three), and the abilities to analyze, synthesize, design, and create (Level Four).

In new assessments, these abilities should be the focus of at least one third of the total points in mathematics and at least half in English language arts. Plans for new consortia assessments intend to meet these benchmarks. This will require a shift from multiple-choice items to more elaborated responses.

2: High-Fidelity Assessment of Critical Abilities

Assessments should evaluate the critical abilities articulated in the standards, such as communication (speaking, reading, writing, and listening in multimedia forms), collaboration, modeling, complex problem solving, research, experimentation, and evaluation.

Figure 2: Depth of Knowledge Levels



Tasks should measure these abilities as they will be used in the real world, rather than through an artificial proxy. For example, while a test might ask a student to point out evidence for a claim in a text, this would not demonstrate that the student knows how to find, evaluate, and use appropriate evidence to build an argument or evaluate a situation. These skills should be directly represented in authentic research tasks.

The new Consortium assessments will measure some of these key abilities, such as listening, writing with revision, and modeling. States and districts will need to augment the tests with performance assessments in order to assess other abilities, such as long-term investigations and spoken, visual, and technology-supported presentations, as many did during the 1990s.

3: Standards that Are Internationally Benchmarked

Assessments should be as rigorous as those of the leading education countries, in terms of the kinds of tasks they present as well as the level of performance they expect.

On the Program in International Student Assessment (PISA) tests, most items require constructed responses to questions that ask students to analyze and apply knowledge to novel problems or contexts. In 2015, PISA will add assessment of collaborative problem-solving to its assessments of reading, mathematics, and scientific literacy. Assessment of computer literacy will follow.

From Finland to Singapore and Australia to New Zealand, students write even more extended responses to questions that require them to evaluate and analyze texts, data, and problems, rather than bubbling in responses to multiple-choice questions. Many examination systems now feature project components that require students to investigate problems and design solutions, conduct research, analyze data, write extended papers, and deliver oral presentations describing their results.

4: Use of Items that Are Instructionally Sensitive and Educationally Valuable

Assessment tasks should also represent the curriculum content in ways that respond to instruction and have value for guiding and informing teaching.

Instructionally sensitive items are designed so that the underlying concepts can be taught and learned, rather than reflecting students' differential access to outside-of-school experiences (frequently associated with their socioeconomic status or cultural context) or depending mostly on test-taking skills. It is not a good use of valuable instructional time to spend hours teaching students to "psych out" the tests rather than to develop the skills they will need to use in the real world.

In many countries, the use of assessments *of, as, and for* learning is a goal. Performance-based assessments are designed to offer good models for teaching and learning and insights into how students think as well as what they know. Assessments that provide these insights, used to guide instruction and revision of work, can be powerful influences on learning and achievement.

5: Assessments that Are Valid, Reliable, and Fair

To be valid for any purpose, an assessment should represent well the knowledge and skills it intends to measure, be used appropriately for intended purposes, and have positive consequences for instruction and for test-takers, guiding better decisions rather than restricting opportunities.

In order to have assessments that are truly valid for a wide range of learners, they should *accurately* evaluate students' abilities and do so *reliably* across testing contexts and scorers. They should also be *fair* and *accessible*: free from bias and designed to reduce unnecessary obstacles to performance that could undermine validity for some subgroups (for example, language complexities not related to the construct being measured).

Use of the principles of universal design, together with the design of accommodations and modifications, should create maximum access to the assessment for a wide range of learners. And they should sufficiently cover the continuum of achievement so that they enable a wide range of students to show what they know and how they've progressed. Finally, they should be transparent enough to support relevant opportunities to learn.

Indicators of Quality in a System of Next Generation Assessments

1: Assessment of Higher-Order Cognitive Skills

- ✓ A large majority of items and tasks (at least two-thirds) evaluate the conceptual knowledge and applied abilities that support transfer (e.g., Depth of Knowledge Levels Two, Three, or Four in Webb's Taxonomy or the equivalent)
- ✓ At least one-third of the assessment content in mathematics, and at least one-half in English language arts, should evaluate higher-order skills that allow students to become independent thinkers and learners (DOK Levels Three or Four)

2: High-Fidelity Assessment of Critical Abilities

Critical abilities outlined in the Standards are evaluated using high-fidelity tasks that use the skills in authentic applications:

- ✓ Research, including analysis and synthesis of information
- ✓ Experimentation and evaluation
- ✓ Oral communications: speaking and listening
- ✓ Written communications: reading and writing
- ✓ Use of technology for accessing, analyzing, and communicating information
- ✓ Collaboration
- ✓ Modeling, design, and problem solving using quantitative tools

3: Standards that Are Internationally Benchmarked

- ✓ Calibration to PISA, International Baccalaureate, or other internationally comparable assessments (based on evaluation of content comparability, performance standards, and analysis of student performance on embedded items)

4: Items that Are Instructionally Sensitive and Educationally Valuable

- ✓ Research that confirms instructional sensitivity
- ✓ Rich feedback on student learning and performance
- ✓ Tasks that reflect and can guide valuable instructional activities

5: Assessments that Are Valid, Reliable, and Fair

- ✓ Evidence that the intended knowledge and skills are well measured
- ✓ Evidence that scores are related to the abilities they are meant to predict
- ✓ Evidence that the assessments are well-designed and valid for each intended use — and that uses are appropriate to the test purposes and validity evidence
- ✓ Evidence that the assessments are unbiased and fairly measure the knowledge and skills of students from different language, cultural, and income backgrounds, as well as students with learning differences
- ✓ Evidence that the assessments measure students learning accurately along a continuum of achievement, consistent with the purposes the assessments are intended to serve

This brief summarizes a full report published by the Stanford Center for Opportunity Policy in Education, the Center for Research on Student Standards and Testing, and the Learning Sciences Research Institute. The full report can be downloaded at: <http://edpolicy.stanford.edu>



**National Center for Research
on Evaluation, Standards, & Student Testing**