

## Today's Tests Seen as Bar to Better Assessment

**Commission says too many tests are used to gauge basic skills**

By **Sarah D. Sparks**

Washington

The use of testing in school accountability systems may hamstring the development of tests that can actually transform teaching and learning, experts from a national assessment commission warn.

Members of the Gordon Commission on the Future of Assessment in Education, speaking at the annual meeting of the National Academy of Education here Nov. 1-3, said that technological innovations may soon allow much more in-depth data collection on students, but that current testing policy calls for the same test to fill too many different and often contradictory roles.

The nation's drive to develop standards-based accountability for schools has led to tests that, "with only few exceptions, systematically overrepresent basic skills and knowledge and omit the complex knowledge and reasoning we are seeking for college and career readiness," the commission writes in one of several [interim reports](#) discussed at the Academy of Education meeting.

"We strongly believe that assessment is a primary component of education, ... [part of] the trifecta of teaching, learning, and testing," said Edmund W. Gordon, the chairman of the commission and a professor emeritus of psychology at Yale University and Teachers College, Columbia University.

The two-year study group launched in 2011 with initial funding from the Princeton, N. J.-based Educational Testing Service and a membership that reads like a [who's who](#) of education research and policy. Its 32 members include: author and education historian Diane Ravitch of New York University, former West Virginia Gov. Bob Wise of the Washington-based Alliance for Excellent Education, and cognitive psychologist Lauren Resnick of the University of Pittsburgh, among others.

The panel is developing recommendations for both research on new assessments—for the Common Core State Standards and others—and policy for educators on how to use tests appropriately. The final recommendations, expected at the end of the year, will be based on two dozen studies and

[← Back to Story](#)

EDUCATION WEEK  
**Multi-User  
LICENSING**



**UNLOCK  
ESSENTIAL  
CONTENT**  
Open your  
team's potential

**+ Click here for more info**

analyses from experts in testing on issues of methods, student privacy, and other topics.

## Stopping Overlap

Education policymakers understandably want to develop multimillion-dollar tests as efficiently as possible, said Lorrie A. Shepard, the education dean at the University of Colorado at Boulder and part of the commission's executive council. However, she said, they often confuse summative tests—large-scale snapshots such as the standardized tests states use for accountability—with formative tests, which are used to diagnose specific learning problems in individual students and improve instruction over time.

"This set of misbeliefs is actually fostering worse and worse tests," which assess only surface details that can be gathered for quick turnaround, rather than more-nuanced measures of deep knowledge, retention, and the ability to transfer knowledge to other subjects, she said.

Because teachers are accountable—and increasingly evaluated professionally—on the basis of those tests, Ms. Shepard added, "the way math and reading are taught are disabling because they are taught for recognition and taught for memorization, and even comprehension is being postponed. The way those subject matters get presented is the harm of those teaching-to-the-test regimes."

Both Ms. Shepard and Elena Silva, a senior policy analyst at Education Sector, a Washington think tank, said commercial testing companies increasingly offer electronic versions of tests that don't gauge deeper learning. Ms. Shepard said that education needs a *Consumer Reports* to identify tests being used for purposes for which they were not designed.

Test developers and policymakers alike should think of tests as a framework to create feedback loops for improvement, argued Robert J. Mislevy, the chairman for measurement and statistics of the ETS and part of the commission's executive council.

"Who needs the information at what time?" Mr. Mislevy said. "Sometimes feedback loops are very tight—when you're playing a learning game, for example, feedback loops are taking place in a second or two. There are other feedback loops that are much bigger, like those used by chief state school officers looking at policy over the course of years."

Those assessments, rather than being used simply to rank students, could help educators identify learning patterns, he said.

For example, the ETS' Global Integrative Summative Assessment, now in field-testing as part of the federal Reading for Understanding program, uses scenarios to differentiate a student's comprehension ability from his or her background knowledge.

Each scenario in the test is a cycle. Students first are tested on vocabulary and concepts related to a topic. Then they read a passage on the topic and summarize the main idea and key details of the text.

Finally, the students report on how they incorporate what they have read into what they already knew about the topic—for example, by completing an interactive graphic, according to Barbara Foorman, an education professor and the director of the Florida Center for Reading Research at Florida State University in Tallahassee. She spoke in an interview with *Education Week*.

### RELATED BLOG



[Visit this blog.](#)

Nuanced test design would not replace the need for separate formative and summative tests, Mr. Mislevy of the ETS said, but it could help educators and policymakers think differently about what can be learned from tests.

Vol. 32, Issue 12, Page 8