

Beyond Basic Skills



**The Role of Performance
Assessment in Achieving 21st
Century Standards of Learning**

Commissioned Papers

Suzanne Lane, *Performance Assessment: The State of the Art.*

Raymond Pecheone and Stuart Kahl, *Developing Performance Assessments: Lessons from the United States.*

Brian Stecher, *Performance Assessment in an Era of Standards-Based Educational Accountability.*

Jamal Abedi, *Performance Assessments for English Language Learners.*

Linda Darling-Hammond, with Laura Wentworth, *Benchmarking Learning Systems: Student Performance Assessment in International Context.*

Lawrence Picus, Will Montague, Frank Adamson, and Maggie Owens, *A New Conceptual Framework for Analyzing the Costs of Performance Assessment.*

Barry Topol, John Olson, and Edward Roeber, *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments.*

Advisory Committee

Chair - Richard Shavelson, Margaret Jacks Professor of Education and Psychology,
Stanford
University

Eva Baker, Professor, UCLA, and Director of the Center for Research on Evaluation,
Standards, and Student Testing

Christopher Cross, Chairman, Cross & Jofus, LLC

Nicholas Donahue, President and CEO, Nellie Mae Education Foundation, and
former
State Superintendent, New Hampshire

Michael Feuer, Executive Director, Division of Behavioral and Social Sciences and
Education in the National Research Council (NRC) of the National Academies

Edward Haertel, Jacks Family Professor of Education, Stanford University

Jack Jennings, President and CEO, Center on Education Policy

Peter McWalters, Strategic Initiative Director, Education Workforce, Council of Chief
States School Officers (CCSSO) and former State Superintendent, Rhode Island

Lorrie Shepard, Dean, School of Education, University of Colorado at Boulder

Guillermo Solano-Flores, Professor of Education, University of Colorado at Boulder

Brenda Welburn, Executive Director, National Association of State Boards of
Education

Gene Wilhoit, Executive Director, Council of Chief States School Officers

A Call for Next Generation Assessments

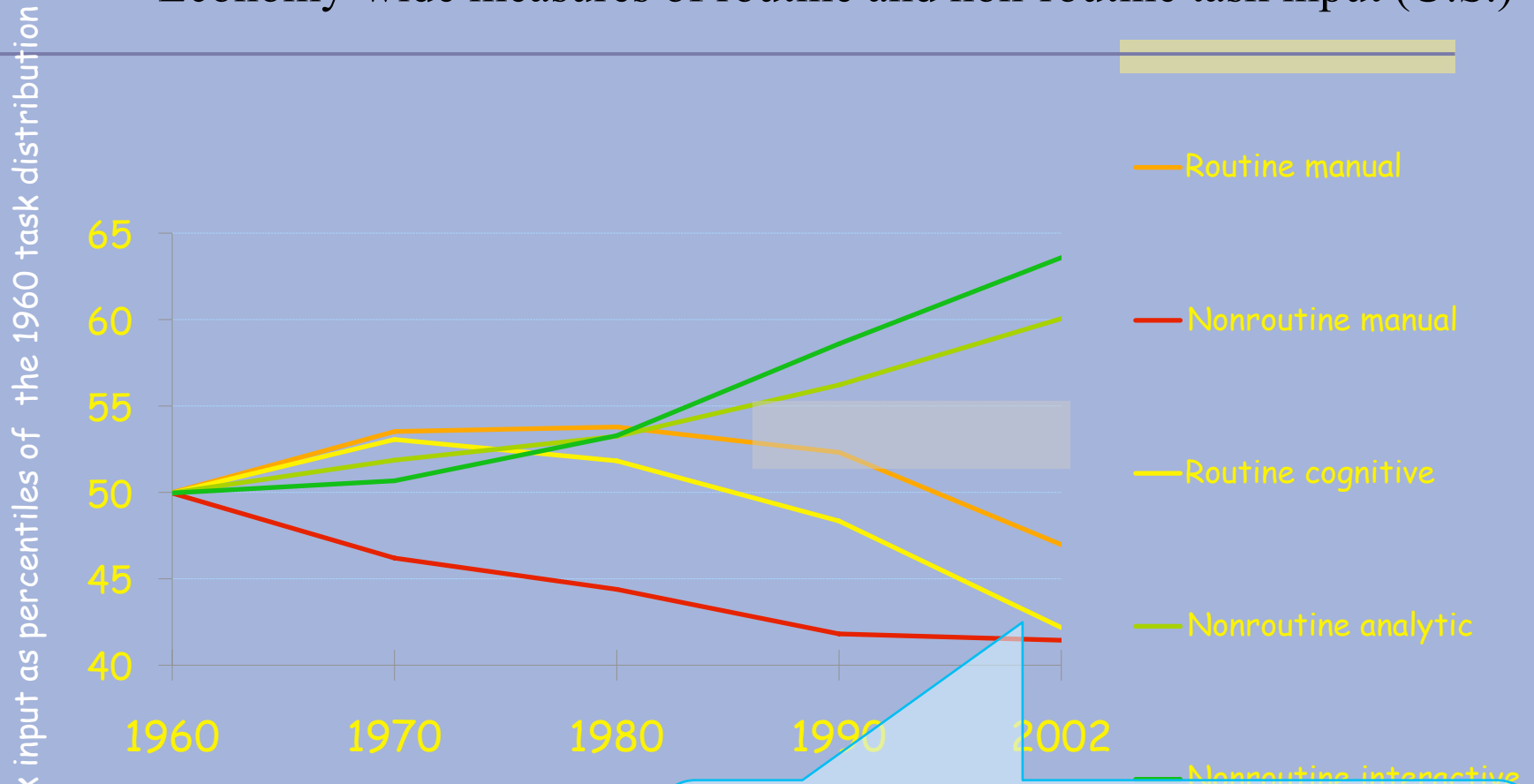
“I am calling on our nation’s Governors and state education chiefs to develop standards and assessments that don’t simply measure whether students can fill in a bubble on a test, but whether they possess 21st century skills like problem-solving and critical thinking, entrepreneurship and creativity.”

-- President Barack Obama

March 10, 2009

How the demand for skills has changed

Economy-wide measures of routine and non-routine task input (U.S.)



(Levy and Murnane)

The dilemma of schools:

The skills that are easiest to teach and test are also the ones that are easiest to digitize, automate, and outsource

Concerns about the Effects of High-Stakes Multiple-Choice Tests on Instruction

“I have seen more students who can pass the [state test] but cannot apply those skills to anything if it’s not in the test format. I have students who can do the test but can’t look up words in a dictionary and understand the different meanings.... As for higher quality teaching, I’m not sure I would call it that. Because of the pressure for passing scores, more and more time is spent practicing the test and putting everything in [the test] format.”

-- A Texas Teacher

High-Achieving Countries Rely Increasingly on Performance Assessments

While multiple choice testing predominates in the U.S.,

To “B” or Not to “B”

most high-

achieving countries largely

use written &



assessments, plus
samples of student work,

such as research projects
and exhibitions, to evaluate
what students have

learned.

Hong Kong's Rationale for Increasing School-Based Assessments

“The primary rationale for School Based Assessments (SBA) is to enhance the validity of the assessment, by including the assessment of outcomes that cannot be readily assessed within the context of a one-off public examination....”

“SBA, which typically involves students in activities such as making oral presentations, developing a portfolio of work, undertaking fieldwork, carrying out an investigation, doing practical laboratory work or completing a design project, help students to acquire important skills, knowledge and work habits that cannot readily be assessed or promoted through paper-and-pencil testing. Not only are they outcomes that are essential to learning within the disciplines, they are also outcomes that are valued by tertiary institutions and by employers.”

(Hong Kong Education Examinations Authority, 2009).

Singapore GCE A-Level Examinations

Time-based Written Papers

- 3 hour duration; 2 to 4 papers per H2 subject
- Open-ended essays, structured questions, case studies, source-based questions
- Externally set and marked by SEAB/CIE

School- based Coursework

- Longer duration of about 6 months
- Product (e.g. Artwork or design task), Oral Presentation, Independent Study
- Tasks set by SEAB/CIE, internally marked by teachers, externally moderated by SEAB/CIE)

SCHOOL-BASED SCIENCE PRACTICAL ASSESSMENT

To Assess Experimental Skills and Investigations, Students...

- Identify a problem, design and plan an investigation, evaluate their methods and techniques
- Follow instructions and use techniques, apparatus and materials safely and effectively
- Make and record observations, measurements, methods, and techniques with precision and accuracy
- Interpret and evaluate observations and experimental data

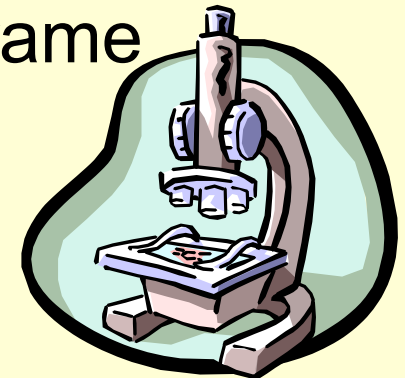


What is Performance Assessment?

Students must construct an answer, produce a product, or perform an activity.

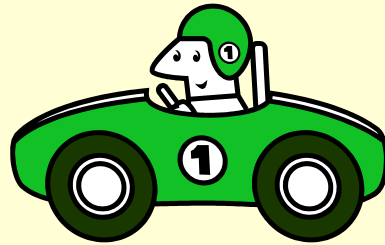
PA may range from a short response or problem solution, to writing an essay, to designing, conducting, and analyzing a laboratory experiment.

PA measures students' reasoning skills and their ability to apply knowledge to frame and solve meaningful problems.



Performance Assessments are Common in Many Arenas

Driving Tests



Eye Exams



Sports Events



Medical Boards



Engineering Exams



Ohio Performance Assessment Project

“Heating Degrees” Task

- Task requires mathematical analysis and modeling as well as sophisticated understanding of ratio and proportion
- Response requires display of mathematical problem-solving, use of on-line research, explanation, and defense of ideas
- Knowledge is applied to a complex, real-world problem

Gas Bills, Heating Degree Days, and Energy Efficiency

Here is a typical story about an Ohio family concerned with saving money and energy by better insulating their house.

Kevin and Shana Johnson's mother was surprised by some very high gas heating bills during the winter months of 2007. To improve the energy efficiency of her house, Ms. Johnson found a contractor who installed new insulation and sealed some of her windows. He charged her \$600 for this work and told her he was pretty sure that her gas bills would go down by "at least 10 percent each year." Since she had spent nearly \$1,500 to keep her house warm the previous winter, she expected her investment would conserve enough energy to save at least \$150 each winter (10% of \$1,500) on her gas bills.

Ms. Johnson's gas bill in January 2007 was \$240. When she got the bill for January 2008, she was stunned that the new bill was \$235. If the new insulation was going to save only \$5 each month, it was going to take a very long time to earn back the \$600 she had spent. So she called the insulation contractor to see if he had an explanation for what might have gone wrong. The contractor pointed out that the month of January had been very cold this year *and* that the rates had gone up from last year. He said her bill was probably at least 10% less than it would have been without the new insulation and window sealing.

Ms. Johnson compared her January bill from 2008 to her January bill from 2007. She found out that she had used 200 units of heat in January of 2007 and was charged \$1.20 per unit (total = \$240). In 2008, she had used 188 units of heat but was charged \$1.25 per unit (total = \$235) because gas prices were higher in 2008. She found out the average temperature in Ohio in January 2007 had been 32.9 degrees, and in January of 2008, the average temperature was more than 4 degrees colder, 28.7 degrees. Ms. Johnson realized she was doing well to have used less energy (188 units versus 200 units), especially in a month when it had been colder than the previous year.

Since she used gas for heating only, Ms. Johnson wanted a better estimate of the savings due to the additional insulation and window sealing. She asked Kevin and Shana to look into whether the "heating degree days" listed on the bill might provide some insight.

Argon Energy Co.		Customer	Bill Date
		Ms. Ariane Johnson 48 Bluebonnet Avenue Columbus OH 43206	Jan 31, 2008
			Account # 56-73342B Residential
Current Itemized Bill			
	December 30 reading actual		8300
	January 31 reading actual		8488
	Total units used January 2008		188
January 2008:	1108 heating degree days 0 cooling degree days		
	Price per unit @ \$1.25		\$235
Energy Use History			
	Total units used January 2007		200
January 2007:	1000 heating degree days 0 cooling degree days		
TOTAL CURRENT CHARGES			\$235

Ohio Performance Assessment Project

“Heating Degrees” Task

Based on Ms. Johnson’s situation and some initial information to begin to research “heating degree days” on the internet:

- (1) **Assess the cost-effectiveness of Ms. Johnson’s new insulation and window sealing.** In your assessment, you must do the following:
 - Compare Ms. Johnson’s gas bills from January 2007 and January 2008.
 - Explain Ms. Johnson’s savings after the insulation and sealing.
 - Identify circumstances under which Ms. Johnson’s January 2008 gas bill would have been at least 10% less than her January 2007 bill.
 - Decide if the insulation and sealing work on Ms Johnson’s house was cost-effective and provide evidence for this decision.

- (2) **Create a short pamphlet for gas company customers to guide them in making decisions about increasing the energy efficiency of their homes.** The pamphlet must do the following:
 - List the quantities that customers need to consider in assessing the cost-effectiveness of energy efficiency measures.
 - Generalize the method of comparison used for Ms. Johnson’s gas bills with a set of formulas, and provide an explanation of the formulas.
 - Explain to gas customers how to weigh the cost of energy efficiency measures with savings on their gas bills.

Why Use Performance Assessments as Part of an Assessment System?

Performance assessments have been found to:

- Develop and evaluate critical thinking and performance abilities
- Provide more information to inform planning and instruction
- Improve the quality of instruction, including the teaching of more challenging knowledge and skills, and the expectation of greater production, explanation & revision of work to standards
- Help teachers and students internalize standards through involvement in scoring and feedback
- Improve preparation for the demands of college and work.

Challenges of Performance Assessment

- **Reliability and Validity**
- **Fairness**
- **Feasibility**
- **Costs**



Reliable and Valid Assessments

Careful Design: Tasks and rubrics are

- Well-linked to standards
- Designed to measure clearly defined knowledge and skills
- Based on a clear understanding of cognitive development
- Expressed in criteria that define a competent performance,
- Rigorously field tested to ensure that the items or tasks are understandable and are measuring the intended concepts and abilities.

When these principles are followed, studies have found that assessments can be made comparable and valid across time, tasks, and raters

Reliable Scoring Systems

- ❏ Are based on standardized tasks
- ❏ Feature well-designed scoring rubrics
- ❏ Carefully train scorers
- ❏ Moderate the scoring process to ensure consistency in applying the standards,
- ❏ Audit the system to double check and upgrade comparability.

Well-developed systems with these features have produced inter-rater reliability with levels of agreement of 90% or higher.

Methods for Ensuring Fairness

- Use universal design principles,
- Make linguistic choices to avoid sources of confusion unrelated to the content being measured
- Review items and tasks for cultural bias
- Pilot test tasks to see how they perform with different test-takers.

Carefully designed performance assessments have been found to produce more successful evaluations of knowledge for English learners, special education students, and students who struggle in other ways than some traditional standardized tests.

Improved Feasibility

- Design enhancements can create more reliably useful and generalizable tasks
(Task shells, uses of learning progressions)
- Development of state capacity
 - Creation of systems for development, scoring, and auditing
 - Integration into PD
- Technological advances



Technology Uses

- ❏ To distribute and administer assessments
- ❏ To enable simulations, research tasks, and other means for evaluating applied learning
- ❏ To support both human scoring and machine scoring of open-ended items

As a measure of the potential for technology to streamline performance testing, the National Assessment of Educational Progress has found that human and computer scoring of a set of physics simulations matches 96 percent of the time.

NAEP 8th Grade Computer-Based Simulation Task

As part of an Investigation of why scientists use helium gas balloons to explore outer space and the atmosphere, students explore a simulated world-wide web:

“Some scientists study space with large helium gas balloons. These balloons are usually launched from the ground into space but can also be launched from a spacecraft near other planets.

Why do scientists use these gas balloons to explore outer space and the atmosphere instead of using satellites, rockets, or other tools? Base your answer on more than one web site.”

After a hands-on task in which students design, conduct, and evaluate a scientific investigation, “How do different amounts of helium affect the altitude of a helium balloon? Support your answer with what you saw when you experimented.”



ASG Cost Comparisons

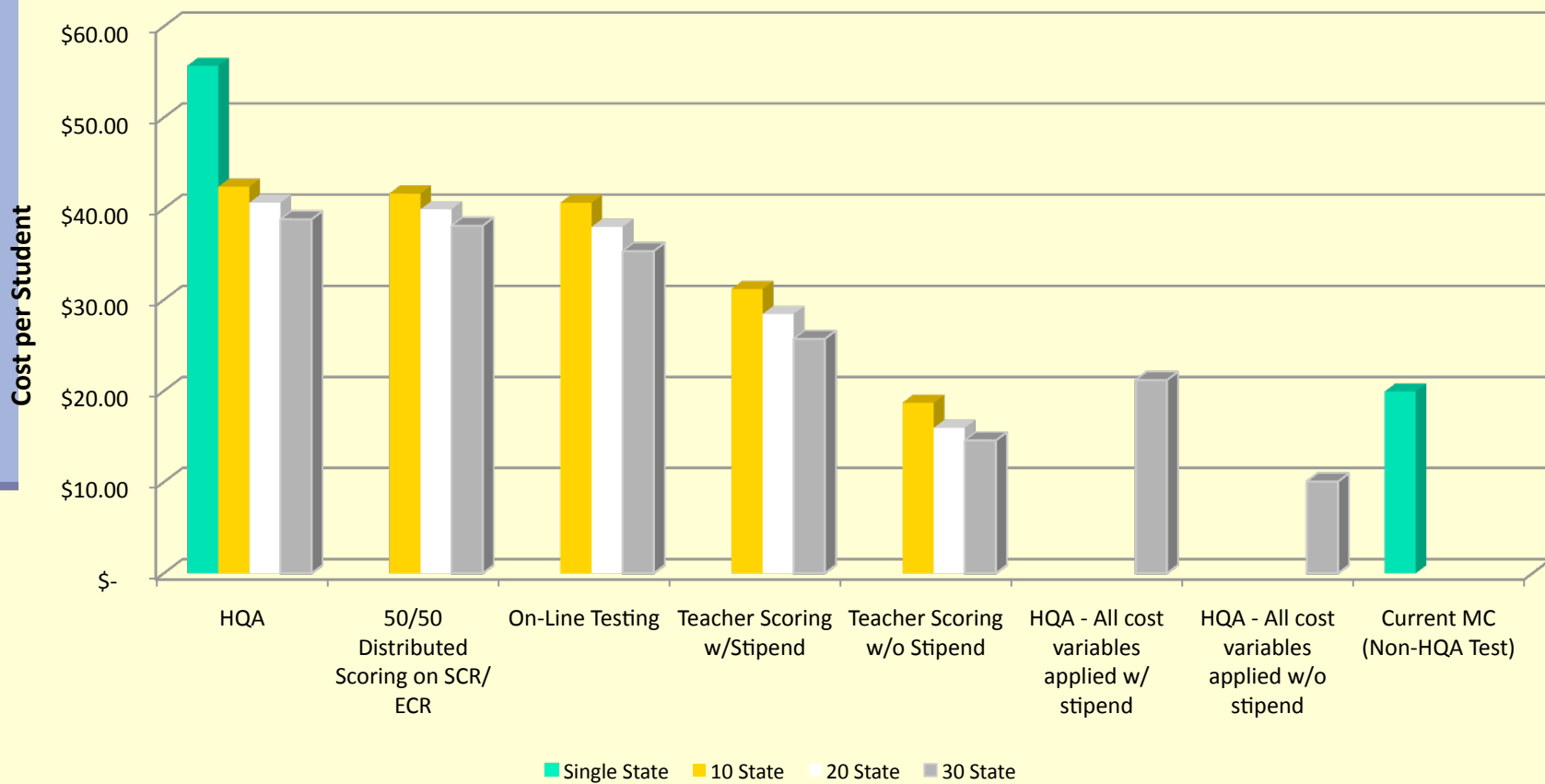
	Item Types	Cost per Student
Typical Assessment Program	50 multiple- choice 1-2 constructed response items	\$20 / student
High-Quality Summative Assessment Program	25 multiple – choice items; 4 constructed response items; 2-3 performance tasks	\$55 / student Assuming a single state implementation and no cost reduction strategies

Cost Savings from

- Developing assessments in Consortia
- Using teachers to score
- Using technology for
 - on-line assessment
 - distributed scoring
 - AI scoring

Overall Cost Reduction Results

Diminishing Expenditures - per Student HQA Cost



Potential Benefits

- Exemplars of transferable knowledge and skills
- More rigorous and cognitively demanding instruction
- Teacher and student internalization of standards
- More useful information about student thinking and performance to inform instructional decision making
- Opportunity to improve use of professional development time and resources

Issues for the Future

- Making thoughtful decisions about what standards are best assessed in what ways
- Ensuring care in design and balance in implementation
- Developing systems for supporting strategic uses of human and machine scoring
- Investing in infrastructure and capacity at the state and local levels

All reports can be
downloaded from
<http://edpolicy.stanford.edu>