

NASSP Bulletin

<http://bul.sagepub.com>

Setting Standards for Students: The Case for Authentic Assessment

Linda Darling-Hammond
NASSP Bulletin 1993; 77; 18
DOI: 10.1177/019263659307755604

The online version of this article can be found at:
<http://bul.sagepub.com/cgi/content/abstract/77/556/18>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:

[National Association of Secondary School Principals](#)

Additional services and information for *NASSP Bulletin* can be found at:

Email Alerts: <http://bul.sagepub.com/cgi/alerts>

Subscriptions: <http://bul.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Setting Standards for Students: The Case for Authentic Assessment

By Linda Darling-Hammond

When assessment allows students to achieve challenging goals in authentic ways, it creates more than just high scores—it creates confident and capable learners.

As students prepare for graduation at Central Park East Secondary School (CPESS), a high school of 450 students in an East Harlem neighborhood in New York City, they do not worry about assembling Carnegie Units, nor do they cram for multiple-choice Regents examinations as do many other students in high schools around New York. Instead, they work intensively during their one to three years in the CPESS Senior Institute preparing a portfolio of their work that will reveal their competence and performance in 14 curricular areas, ranging from science and technology to ethics and social issues, from school and community service to mathematics, literature, and history.

This portfolio will be evaluated by a graduation committee composed of teachers from different subjects and grade levels, an outside examiner, and a student peer. The committee members

Linda Darling-Hammond is at Teachers College, Columbia University, New York City.

will examine all the entries and hear the students' oral "defense" of their work as they determine when each student is ready to graduate.

Part of the growing movement to establish means for more authentic assessment of student learning, CPESS is developing ways to focus students' energies on challenging performance-oriented tasks that require analysis, integration of knowledge, and invention as well as highly developed written and oral expression rather than merely recall and recognition of facts. Students develop a project that demonstrates their knowledge of scientific methodology, for example, by using it in a particular field. They engage ethical and social issues by participating in a debate, writing an op-ed article, or analyzing a film or novel that raises important moral issues.

In each case, they must demonstrate their ability to see multiple viewpoints, weigh conflicting claims, and defend their views with credible evidence. Literary essays and historical analyses, along with documentation and evaluation of their internship experiences, add to the wealth of evidence students' accumulate about their attainment of valued school—and societal—goals.

Increasingly, local schools, districts, and states are experimenting with these methods and other alternatives to standardized testing for assessing student learning and performance. Much like the kinds of assessments that prevail in most other countries around the world—where multiple-choice testing is

much less common—these approaches include essay examinations, research projects, scientific experiments, oral exhibitions, and performances in such areas as debating and the arts.

They also include portfolios of students' work in various subject areas, along with individual and group projects requiring analysis, investigation, experimentation, cooperation, and written, oral, or graphic presentation of findings. Often, the assessment occasion requires students to respond to questions from classmates or from external examiners, thus helping them learn to think through and defend their views, while allowing their teachers to hear and understand their thinking (Coalition of Essential Schools, 1990; Archbald and Newmann, 1988).

Why Alternative Assessment?

One of the reasons for these efforts to develop alternative forms of assessment is a growing consensus among educators, researchers, and policymakers that current U.S. tests do not tap many of the skills and abilities that students need to develop in order to be successful in later life and schooling. These concerns are partly due to the limits of widely used U.S. testing methods, described more fully below. The concerns are also related to the increasing demands for a kind of education that encourages students to do more than memorize information and use it to solve tidy problems—an education

that prepares students to frame problems, find information, evaluate alternatives, create ideas and products, and invent new answers to messy dilemmas.

The capacities required of students today are more demanding than those required in the past. A growing number of jobs in our information economy require highly developed intellectual skills and technological training. Even "low-skill" jobs require technical training and flexibility. In addition, most industries are restructuring the way they organize work so that cooperative planning and problem solving are the "basic skills" that have replaced following simple directions on an assembly line.

Citizens must be able to access resources and perform complicated tasks at high levels of literacy to survive in today's world. Workers must anticipate changing occupations several times over the course of a lifetime, adapting to ever-changing technologies and job demands, and inventing solutions to productivity problems rather than relying on a manager to tell them what to do (Hudson Institute, 1987; Drucker, 1986).

These kinds of skills and abilities are not based on the kinds of thinking and performance that are evaluated in most U.S. testing programs. Because of the way widely used multiple-choice, norm-referenced tests are constructed, they exclude a great many kinds of knowledge and types of performance we expect from students, placing testtakers in a pas-

sive, reactive role, rather than one that engages their capacities to structure tasks, produce ideas, and solve problems (National Research Council, 1982). Current research on human learning and performance has suggested that many currently used tests fail to measure students' higher order cognitive abilities or to support their capacities to perform real world tasks (Resnick, 1987; Sternberg, 1985).

Concerns About Current Testing Policies

These shortcomings were less problematic when they were used as only one of many kinds of information about student learning, and when they were not directly tied to decisions about students and programs. However, as test scores have been increasingly used to make important educational decisions, their flaws have become more damaging.

As schools have begun to "teach to the tests," the scores have become ever poorer assessments of students' overall abilities, because classwork oriented toward recognizing the answers to multiple-choice questions does not heighten students' proficiency in areas that are not tested, such as analysis, complex problem solving, and written and oral expression (Koretz, 1988; Haney and Madaus, 1986; Darling-Hammond and Wise, 1985).

Because teachers must emphasize those things that tests measure, current approaches to testing often limit the kinds of teaching and learning

opportunities provided in classrooms.

The results of this phenomenon can be seen in U.S. students' performance. Since about 1970, when standardized tests began to be used for a wider variety of accountability purposes, basic skills test scores have been increasing slightly, while assessments of higher order thinking skills have declined in virtually all subject areas. Officials of the National Assessment of Educational Progress, the National Research Council, and the National Councils of Teachers of English and Mathematics, among others, have all attributed this decline in higher order thinking and performance to schools' emphasis on tests of basic skills. They argue that not only are the test scores inadequate measures of students' performance abilities, but also that the uses of the tests have corrupted teaching practices.

As the National Assessment of Educational Progress found: "Only 5 to 10 percent of (high school) students can move beyond initial readings of a test; most seem genuinely puzzled at requests to explain or defend their points of view." The NAEP assessors explained that current methods of teaching and testing reading require short responses and lower level cognitive thinking, resulting in "an emphasis on shallow and superficial opinions at the expense of reasoned and disciplined thought..., (thus) it is not surprising that students fail to develop more comprehensive thinking and analytical skills" (NAEP, 1981).

A more recent NAEP report summarized the status of high school students' performance as follows:

Sixty-one percent of the 17-year-old students could not read or understand relatively complicated material, such as that typically presented at the high school level. Nearly one-half appear to have limited mathematics skills and abilities that go little beyond adding, subtracting, and multiplying with whole numbers. More than one-half could not evaluate the procedures or results of a scientific study, and few included enough information in their written pieces to communicate their ideas effectively. Additionally, assessment results in other curriculum areas indicate that high school juniors have little sense of historical chronology, have not read much literature, and tend to be unfamiliar with the uses and potential applications of computers (ETS, 1989, p. 26).

International comparisons of student performance in mathematics and science tell a similar story. U.S. students score at about the median of other countries at fifth grade, dip below the average by eighth grade, and consistently score near the bottom by twelfth grade, especially on tasks requiring higher order thinking and problem solving.

An international mathematics study found that, in line with U.S. testing demands, instruction in the United States is dominated by textbooks and lectures, followed by individual seatwork, with little use of other resources such as computers,

calculators, or manipulatives.

The researchers concluded that these “strategies geared to rote learning” represent:

... a view that learning for most students should be passive—teachers transmit knowledge to students who receive it and remember it mostly in the form in which it was transmitted.... In the light of this, it is hardly surprising that the achievement test items on which U.S. students most often showed relatively greater growth were those most suited to performance of rote procedures (McKnight et al., 1987, p. 81).

Two recent major studies called attention to this problem. Ernest Boyer’s (1983) study of U.S. high schools found an overabundance of teaching consisting of the transmittal of “fragments of information, unexamined and unanalyzed.”

Boyer notes:

The pressure is on to teach the skills that can be counted and reported. As one teacher said, “We are so hung up on reporting measured gains to the community on nationally normed tests that we ignore teaching those areas where it can’t be done.”

Similarly, John Goodlad (1984) found in his massive study of more than 1,000 U.S. classrooms that for the most part, “the curriculum appeared to call for and make appropriate only some ways of knowing and learning and not others.” He found that students listen, read short sections in textbooks,

respond briefly to questions, and take short-answer or multiple-choice quizzes. They rarely plan or initiate anything, create their own products, read or write anything substantial, or engage in analytic discussions. And there are few incentives for their teachers to pursue these approaches.

As Goodlad comments:

Teachers are sensitive to the pressures that state and district testing programs place on them. They get the message. The other messages—that there are goals beyond those that the tests measure, that pursuing such goals calls for alternative teaching strategies, that the fundamentals of the curriculum transcend grade-level requirements—are faint to begin with, and they are drowned out by the more immediate and stronger message...

As a recent study of the implementation of California’s new mathematics curriculum points out, when a curriculum reform aimed at problem solving and the development of higher order thinking skills encounters an already mandated rote-oriented basic skills testing program, the tests win out (Cohen et al., 1990; Darling-Hammond, 1990).

As one teacher put it:

Teaching for understanding is what we are supposed to be doing... (but) the bottom line here is that all they really want to know is how are these kids doing on the tests?...They want me to teach in a way that they can’t test, except that I’m held accountable to the test. It’s a Catch 22.... (Wilson, 1990).

These studies point out how important it is for schools to choose their “accountability tools” carefully. Clearly, if performance measures are actually to support meaningful learning, they must assess and encourage valuable kinds of teaching in classrooms.

What Are the Alternatives?

Assessment in most other countries is substantially different from the kind of multiple-choice testing common in the United States. Not unlike the Advanced Placement tests taken by a small minority of U.S. seniors, high school students in most European countries complete extended essay examinations, often coupled with oral examinations, in a range of subjects requiring serious critical thought. The French Baccalaureate, for example, asks such questions as: “What is judgment?” and, “Why should we defend the weak?” These are a far cry from the kind of thinking required of most U.S. students, who use number two pencil to fill in fixed-response bubbles aimed at identifying a single right answer.

Other countries’ assessments also include practical performance events requiring students to plan, implement, and/or evaluate various tasks, such as the use of scientific procedures or the conduct of a social research project.

Some assessments involve the guided development of cumulative portfolios of student work which shape learning opportunities and classroom evaluation over the course

of a year or more. Graduates in England submit such portfolios, along with written examinations in three of their chosen areas of specialty. The other exhibitions and oral examinations in which they participate are designed to provide many and varied opportunities for them to display their best work, while allowing their teachers and outside examiners opportunities to probe the nature and quality of their thinking.

In most of these countries, there is also a different notion of the role of educators in assessment. Faculties convene to develop and score the assessments. Teachers are involved in examining their own students and those of teachers in other schools. In many cases, much of the assessment process is internal, in the sense that it is under the control of the teacher and directly tied to ongoing instruction. In these ways, the act of assessment improves knowledge, practice, and shared standards across the educational enterprise as a whole, among the professional faculty and the students.

What separates these assessment strategies from the forms of testing more traditional in the United States? According to Wiggins (1989), authentic tests have four basic characteristics in common. First, they are designed to be truly representative of performance in the field. Students actually *do* writing—for real audiences—rather than taking spelling tests or answering questions about writing. They *conduct* science experiments rather than

memorizing disconnected facts about science. The tasks are contextualized, complex intellectual challenges involving the student's own research or use of knowledge in "ill-structured" tasks requiring the development and use of meta-cognitive skills. They also allow appropriate room for student learning styles, aptitudes, and interests to serve as a source for developing competence and for the identification of (perhaps previously hidden) strengths.

Second, the criteria used in the assessment seek to evaluate "essentials" of performance against well-articulated performance standards that are openly expressed to students and others in the learning community, rather than kept secret in the tradition of content-based examinations. These criteria represent a standard because they are based on explicit and shared school-wide aims, and they are multifaceted, representing the various aspects of a task, rather than reduced to a single grade.

Because the criteria are performance-oriented (e.g., demonstrated ability to evaluate competing viewpoints and evidence in developing a persuasive essay concerning a topic of social importance), they guide teaching, learning, and evaluation in a way that illuminates the goals and processes of learning, placing teachers in the role of coach and students in the role of performers as well as self-evaluators.

As suggested above, the third characteristic is that self-assess-

ment plays an important role in authentic tasks. A major goal of authentic assessment is to help students develop the capacity to evaluate their own work against public standards; to revise, modify, and redirect their energies, taking initiative to assess their own progress.

This is a major aspect of self-directed work and self-motivated improvement required of all human beings in real-world situations. Because performance standards take the concept of progress seriously—making the processes of refinement and improvement of products a central aspect of the task and its evaluation—they also allow students of all initial levels of developed competence the opportunity to see, acknowledge, and receive credit for their own growth.

Finally, the students are generally expected to present their work and defend themselves publicly and orally to ensure that their apparent mastery is genuine. This characteristic of authentic assessment serves other goals as well—signaling to students that their work is important enough to be a source of public learning and celebration; providing opportunities for others in the learning community—students, faculty, and parents—to continually examine, refine, learn from, and appreciate shared goals and achievements; and creating living representations of the purposes and standards of the learning community so that they remain vital and energizing.

A number of schools, including

members of the Coalition of Essential Schools founded by TheodoreSizer (1984; 1992), are engaged in creating authentic assessments of student learning. A growing number of states—including Vermont, California, Connecticut, and New York—are developing new approaches to assessment that will transform statewide testing. Teachers in Vermont are developing student portfolios in writing and mathematics as the basis of their state's assessment system.

Connecticut and New York have begun to develop performance-based assessments that require students to perform a science experiment or solve a real-world problem using mathematical and scientific concepts rather than complete a multiple-choice test. California, Maryland, and several other states have developed writing assessments based on student essays. Some of these engage students in complex writing tasks requiring several days of work, including revisions, as part of the examination process. District such as Shoreham-Wading River, N.Y.; Pittsburgh, Pa.; New York City and Rochester, N.Y.; and Albuquerque, N.Mex., are also creating authentic assessments to take the place of standardized testing.

Initiatives such as these are an attempt to make schools genuinely accountable for helping students to acquire the kinds of higher order skills and abilities they will need to use in the world outside school. As the Coalition of Essential Schools (1990) explains:

Of course we want students who are

curious, who know how to approach new problems, who use reading and writing across the disciplines as a natural part of that process, who are thoughtful, able, and active citizens. And to get them we (should) make those goals known from the start, test for them regularly, and correct a student's course when necessary.

In addition to helping teachers and students to evaluate what the students can really do, these approaches serve as expressive tools for students and are highly motivating. Sizer points out that they are as much inspiration as measurement: "Giving kids a really good target is the best way to teach them... And if the goal is cast in an interesting way, you greatly increase the chances of their achieving it" (Coalition of Essential Schools, 1990).

As a testament to Sizer's claim, every one of the 1991 graduates of Central Park East Secondary School went on to postsecondary education, and 92 percent of them were accepted to four-year colleges, a rate more than twice as high as surrounding area high schools.

One student explained their success in terms of the authenticity of goals set and performances achieved: "This environment gives us standards. It makes us look at ourselves in the mirror and feel proud of our accomplishments."

References

Archbald, Doug A., and Newman, Fred M. *Beyond Standardized Testing*:

- Assessing Authentic Academic Achievement in the Secondary School*. Reston, Va.: National Association of Secondary School Principals, 1988.
- Boyer, Ernest. *High School*. New York: Harper and Row, 1983.
- Coalition of Essential Schools. "Performances and Exhibitions: The Demonstration of Mastery." *Horace*, March 1990.
- Cohen, David, et al. "Case Studies of Curriculum Implementation." *Educational Evaluation and Policy Analysis*, Fall 1990.
- Darling-Hammond, Linda. "Instructional Policy into Practice: The Power of the Bottom Over the Top." *Educational Evaluation and Policy Analysis*, Fall 1990.
- Darling-Hammond, Linda, and Wise, Arthur E. "Beyond Standardization: State Standards and School Improvement." *The Elementary School Journal* 3(1985): 315-36.
- Drucker, Peter F. *The Frontiers of Management*. New York: Harper and Row, 1986.
- Educational Testing Service. *A World of Differences: An International Assessment of Mathematics and Science*. Princeton, N.J.: ETS, 1986.
- Goodlad, John. *A Place Called School: Prospects for the Future*. New York: McGraw-Hill, 1984.
- Hudson Institute. *Workforce 2000: Work and Workers for the 21st Century*. Indianapolis, Ind.: Author, 1987.
- Haney, Walt, and Madaus, George. "Effects of Standardized Testing and the Future of the National Assessment of Educational Progress." Working paper for the NAEP study group. Chestnut Hill, Mass.: Center for the Study of Testing, Evaluation and Educational Policy, 1986.
- Koretz, Daniel. "Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction?" *American Educator*, 2(1988): 8-15, 46-52.
- McKnight, C. C.; Crosswhite, F. J.; Dossey, J. A.; Kifer, E.; Swafford, S. O.; Travers, K. J.; and Cooney, T. J. *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective*. Champaign, Ill.: Stipes Publishing, 1987.
- National Assessment for Educational Progress. *Reading, Thinking, and Writing: Results from the 1979-80 National Assessment of Reading and Literature*. Denver, Colo.: NAEP, 1981.
- National Research Council. *Ability Testing: Uses, Consequences, and Controversies*. Edited by Alexandra K. Wigdor and Wendell R. Garner. Washington, D.C.: National Academy Press, 1982.
- Resnick, Lauren B. *Education and Learning to Think*. Washington, D.C.: National Academy Press, 1987.
- Sizer, Theodore. *Horace's Compromise: The Dilemma of the American High School*. Boston, Mass.: Houghton Mifflin, 1984.
- . *Horace's School*. Boston, Mass.: Houghton Mifflin, 1992.
- Sternberg, Robert J. *Beyond IQ*. New York: Cambridge University Press, 1985.
- Wiggins, Grant. "Teaching to the (Authentic) Test." *Educational Leadership*, April 1989.
- Wilson, Suzanne. "A Conflict of Interests: Constraints That Affect Teaching and Change." *Educational Evaluation and Policy Analysis*, Fall 1990.