

Education Policy Analysis Archives

Volume 6 Number 14

July 21, 1998

ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University, Tempe AZ 85287-2411

Copyright 1998, the EDUCATION POLICY ANALYSIS
ARCHIVES. Permission is hereby granted to copy any article
provided that EDUCATION POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

Some Comments on Assessment in U.S. Education

Robert Stake
University of Illinois

Abstract We do not know much about what assessment has accomplished but we know it has not brought about the reform of American Education. The costs and benefits of large scale mandated achievement testing are too complex to be persuasively reported. Therefore, educational policy needs to be based more on deliberated interpretations of assessment, experience, and ideology. Evaluation of assessment consequences, however inconclusive, has an important role to play in the deliberations.

During the last half of the Twentieth Century in America, the traditional quality control of schooling, i.e., informal management (by teachers as well as administrators) board oversight, parent complaint, state guideline and regional accreditation, have continued to be prominent in school operations. But because the perceived quality of public education has fallen off, other means have been added to evaluate and to improve teaching and learning. For thirty years, assessment has been a significant means of quality control and instrument of educational reform.

Earlier, in the Century's third quarter, the impetus for changing American schooling was the appearance of Sputnik. It was reasoned that American schools were unsuccessful if the Soviets could be first to launch spacecraft. College professors and the National Science Foundation stepped forward to redefine mathematics education and the rest of the curriculum, creating a "new math," inquiry teaching, and many courses strange to the taste of most teachers and parents. According to Gallup polls year after year, citizens expressed confidence in the local school but increasingly worried about the national system. In the 1960s, curriculum redevelopment was the main instrument of

reform but, in the 1970s, state-level politicians, reading the public as unhappy both with tradition and federalized reform, created a reform of their own. Their reform spotlighted assessment of student performance.

The term "assessment" then became taken to mean the testing of student achievement with standardized instruments. Student performance goals were made more explicit so that testing could be more precisely focused, and efforts were made to align curricula with the testing. Schooling includes many performances, provisions, and relationships which could be assessed but attention came down predominantly on the students: "If they haven't learned, they haven't been taught."

Now for at least two decades, in almost every school, at every grade level and in each of the subject matters, student achievement has been assessed. And every year, it has been found largely unchanged from previous testing. Over the same periods, teaching, on the whole, appears to have been little changed, certainly not restructured. Explication of goals appears not to have set more achievable targets. The last decade has seen efforts to set standards particularly for levels of student performance needed to restore American Education to a leading, world position. From time to time, gains occurred, but small and not sustained--losses also occurred. Instead of reading this lack of sustained progress as pointing to need for a different grand strategy, the clearest summons has been for additional assessment.

Purposes and Expectations of Assessment

Goal statements are simplifications. The felt purposes of education, aggregated across the profession, across researchers, the public and the primary beneficiaries, are far more complex than those represented in goal statements and formal assessments. Facts, theories, and reasoning are needed not just in isolation but interactively, innovatively, in a range of contexts. We hold a vast inventory of expectations, beyond catalogue, partly ineffable, often only apparent in disappointments as students fall short. That immense inventory is approximated by the informal assessments by teachers much better than by explicated lists of goals.

The grand manifold of purposes of Education held by any one person at any one time also is complex, and situational and internally contradictory. People, even those specially trained, are not very good at speaking of "what all they expect" of an educated person. Again, the complexity shows most forcefully when the person does not perform well. Any one shortfall tells little about the array of purposes. Any one assessment, however precise and valid, does not sample well the manifold of purposes. Broad and attentive use of assessments, formal and informal, evokes realization that what we expect of students and the uses to be made of a graduate's education extend far beyond formal goals, standards and lesson plans. Formal representations of aim and accomplishment provide flimsy accounts of the real thing.

This is not to suggest it useless to record educational purposes and student performance. It is useful to categorize them, to illustrate and prioritize them, sometimes by abilities and subject matters--but always a risk. The subsets or domains are artificial. Needed in the anticipation and provision of Education, they often serve poorly to represent the education a student is attaining. Assessment based strongly on goals or domains is likely to tell more about the territory of teaching than the territory of learning.

Procedurally, Education is organized at the level of courses and classrooms, then lessons and assessments. Actually, education occurs in complex and differentiated ways

in each child's mind. Assessments tuned to management levels cannot be expected to mirror the complexity of learning and diversity of learners. However carefully named and designed, mean scores do not necessarily indicate basic accomplishments for a group of learners. Each testing needs empirical validation.

Validation of Assessment

Standardized test development is one of the most technically sophisticated specialties within Education. Definitions and analytic procedures, at least at the major testing companies are scrutinized, verified, codified and reworked. The traditional ethics of psychometrics call for extensive construct validation of the measurements to be used in schooling. And it is not enough that the instruments and operations be examined for accuracy, relevance and freedom from bias, but that independent measurements be used to confirm that scores indicate what we think they indicate. Sound test development is a slow and expensive procedure.

In the development of assessment instruments by the 50 states, adequate validation has seldom taken place. Instruments have been analyzed statistically to see that they are internally consistent but not that mean what users think they mean. Presumption that assessments indicate quality of teaching, appropriateness of curricula, and progress of the reform movement-- commonplace presumptions in political and media dialogue--is unwarranted. Proper validation would tell us the strength or weakness of our conclusions about student accomplishment. Those studies have not been commissioned. The most needed validation of statewide assessment programs has not taken place.

The question of whether or not the assessment legislation, as opposed to the assessment scores, is having a good effect on student education is a separate question. Assessment changes instruction. Reformists expect assessment will force teachers to teach differently, and, in various ways and to various extents, they do. Each assessment effort will have both positive and negative consequences. The design and promulgation of an assessment program is only an approximation of what actually occurs. The operation described in any report is a partial misrepresentation of institutional initiative and measurement integrity. For a reader, it is an opportunity to misperceive what is happening in the schools and the lives of youngsters. We need better descriptions, better evidence, of those consequences of assessment. And partly because we construct nuances of meaning faster than we invent measurements, we need to understand that we will never have a clear enough picture of the consequences of assessment. All findings should be treated as partial and tentative.

Value Determination

Not only has there been an increase in the amount of formal educational assessment but assessment has been applied increasingly to influence the well-being of students, schools and systems. The "stakes" have risen. Funding, autonomy and privilege have been attached to levels of scoring. The intention has been to get students and teachers dedicated to their tasks, and this sometimes happens, but there have been costs as well as benefits. Among the reported negative consequences of raising the stakes of assessment are:

- instruction is diverted,
- student self-esteem is eroded,
- teachers are intimidated,
- the locus of control of education is more centralized,
- undue stigma is affixed to the school,
- school people are lured towards falsification of scores,
- some blame for poor instruction is redirected toward students when it should rest with the profession and the authorities, and
- the withholding of needed funding for education appears warranted.

The most obvious consequence of increased assessment is that teachers increase preparation for test taking, including test-taking skills and greater familiarization with the anticipated content of testing. Also, topics tested are considered of higher priority and topics untested slip in priority. Assessments are not diagnostic. There is little strategic theory fitting pedagogy to assessment so that few teachers know how to respond to poor student performance, other than to try harder. Thus, over-emphasis on assessment erodes confidence in legitimate teaching competence.

As the stakes rise, the central authorities are both pressured and authorized to intervene more in teaching responsibilities. A widespread public perception of legislators and school authorities is that they are not knowledgeable or competent in matters of the classroom. With ever-confirming evidence that students continue to be testing poorly, the public is tempted to withhold funds for needed improvement in instruction. There is good evidence that increased funding alone will not greatly change the quality of teaching. But at the same time, by investing in the assessment of students without investing in more direct evaluation of teacher and administrative performance, the professional people and the elected overseers are partly "off the hook." In summary, the consequences of assessment are complex, extending far beyond the redirecting of instruction toward state goals.

It is too much to expect that we soon will clearly discern the consequences of assessment and, even less soon, what caused them. Both the consequences and the causes are complex, both as to constituents and as to conditions. Lacking an adequate research base, curricular policy needs to be based on deliberations, long and studied interpretation of assessment, experience, and ideology. That is unlikely when professional wisdom is getting little respect. Often the public presumes that educators put their own interests above those of students. But good deliberations are not uncommon. Evaluation of the consequences of assessment has an important role informing those deliberations.

Even if we were able to improve determination of the consequences of assessment, we lack theory and management systems that guide us in applying that information to the improvement of teaching and learning. We need not wait for politics or the professional to be reformed. We can rely on the political, intuitive, and leadership processes we now have to make assessment more a positive and less a negative force within education.

As indicated before, people do have different purposes for education and for assessment. And for any one purpose, they value the results differently. That is just part of the reality, neither excusing nor facilitating the assessment of assessment.

The assessment practice that does the most measurable, immediate good is not necessarily the practice that has the best long range effect. For example, using testing

time entirely for easily measured skills instead of partly for "ill-defined" interpretive experience increases precision and predictive validity but discourages well-thought-out advocacies to include problem-solving experience throughout elementary school. Value trade-offs need to be considered for long-term as well as short-term effects.

Curriculum and Instruction

Management of teaching and the curriculum cannot be effective without assessment. The best and the worst assessment we have is informal and teacher-driven, sometimes capricious and sometimes more aimed at avoiding embarrassment than maximizing services to children. Yet, it works pretty well, sensitive to what individual children are doing, viewed favorably by a substantial proportion of parents and citizens, especially those people who interact themselves, even in small ways, with the academic program. Still, instructional assessment could be much, much better, and too little professional development is so aimed. The present informal assessment system is little engaged with the formal management information system of school districts and even less with the state's student achievement testing apparatus.

The most successful school improvement efforts have been those that decentralize and protect authority so that a match can be made between what the teachers want to teach and the parents and immediate community want taught. The present decade's "standards movement" was a step in the wrong direction, a further imposition of external values. Assessment was used to nullify decentralization efforts. The state does have a stake in what every child is learning but the state is poorly served by having each child trying to learn the same things. Accountability of the schools is in no way dependent on having each child tied to a core curriculum and tested on the same items. A single test for all is cheaper, but not a service to a diverse population of children.

State assessment is not wrong in its most general finding that teaching and learning in the American schools are mediocre. And that the range across districts is huge. The spread of achievement scores is stable and predictable, more a function of a child's lifetime educational opportunity than of what happens during a year in a classroom. Neither massive changes at home or in the classroom are likely to result in substantial gains on current assessment instruments.

As stated earlier, the validity of measurement of achievement is not the same as validity of those same scores as an indicator of quality of teaching and learning conditions. Teaching can be changed in a number of important ways within a school or classroom without change in achievement means. Using those scores as a measure of school improvement has not been validated. No accumulation of evidence shows assessment to be an indicator of good schooling. In spite of the absence of validity, assessment means continue to be the primary criterion for reform in a vast number of school districts. Given vigorous school improvement efforts over 20-30 years within countless districts, essentially all of them unaccompanied by substantial change in assessment results, what should be concluded is that testing is insensitive to important changes in teaching or that schools cannot be improved. The latter is untenable.

Uses and Stakes

The uses to which assessment information will be put varies not just across assessment approaches but greatly within approaches as well. Different school systems,

teachers, and children, even those greatly alike, will be affected differently. It is not reasonable to suppose that the stakes of assessment are unimportant if they have little impact upon the majority. Special attention needs to be given to how assessment consequences affect the least privileged families and most vulnerable children.

One of the primary stakes of testing is the well-being of teachers. Teachers have much to lose in a high stakes assessment system. Assessment should not be avoided just because teachers protest but their working conditions and professional wisdom should not be trivialized. Teaching quality should be scrutinized. Student performance should be considered but it should not be a primary determinant of teaching competence. There is only a small connection between how well a teacher teaches and how well a child performs on a test.

One of the consequences of high stakes testing is the manipulation of rosters to excuse poor scoring children from participation. The most common way at present appears to be to have children classified as "special education" students, but a good bit of ingenuity has been shown in optimizing rosters.

High stakes assessment often does result in raised scores but the validity of widespread gains, locally or across the country, has not been established. No one wants to challenge the gains that appear, but presently emphasis on small changes serves to orient the school to the assessments rather than to education. Many of the consequences of assessment are best learned from the people who administer the tests, even though they have a self-interest. Many are quick to acknowledge that the assessment enterprise is flawed.

Good research can help but it is mostly a professional and political matter. Until community attitude sets out to make the best of the schools, less to blame them, (however much they deserve the blame), not much good will happen. This is not a nation dedicated to the best possible education system. There are lots of people who would rather have lower taxes than to extend educational benefits. Higher taxes do not assure better opportunities but an interest in finding better opportunities is not a national purpose. Looking at it simplistically, support for assessments appears to be a step toward improving education, but the quarter-century record shows that assessment-driven reform has not worked. Why does it continue to be politically popular? The main consequence of assessment-based reform is that education has not substantially improved. We do not lack evidence of that.

About the Author

Robert E. Stake
University of Illinois--Urbana, Champaign

Email: r-stake@uiuc.edu

Robert Stake is professor of education and director of CIRCE at the University of Illinois. Since 1963 he has been a specialist in the evaluation of educational programs, moving from psychometric to qualitative inquiries. Among the evaluative studies he has directed are works in science and mathematics in elementary and secondary schools, model programs and conventional teaching of the arts in schools, development of teaching with sensitivity to gender equity; education of teachers for the deaf and for

youth in transition from school to work settings, environmental education and special education programs for gifted students, and the reform of urban education. Stake has authored *Quieting Reform*, a book on Charles Murray's evaluation of Cities-in-Schools; two books on methodology, *Evaluating the Arts in Education* and *The Art of Case Study Research*; and *Custom and Cherishing*, a book with Liora Bresler and Linda Mabry on teaching the arts in ordinary elementary school classrooms in America. Recently he led a multi-year evaluation study of the Chicago Teachers Academy for Mathematics and Science. For his evaluation work, in 1988, he received the Lazarsfeld Award from the American Evaluation Association, and, in 1994, an honorary doctorate from the University of Uppsala.

Copyright 1998 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is <http://olam.ed.asu.edu/epaa>

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu. The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu.

EPAA Editorial Board

[Michael W. Apple](#)

University of Wisconsin

[John Covalleskie](#)

Northern Michigan University

[Alan Davis](#)

University of Colorado, Denver

[Mark E. Fetler](#)

California Commission on Teacher Credentialing

[Thomas F. Green](#)

Syracuse University

[Arlen Gullickson](#)

Western Michigan University

[Aimee Howley](#)

Marshall University

[William Hunter](#)

University of Calgary

[Daniel Kallós](#)

Umeå University

[Thomas Mauhs-Pugh](#)

Rocky Mountain College

[Greg Camilli](#)

Rutgers University

[Andrew Coulson](#)

a_coulson@msn.com

[Sherman Dorn](#)

University of South Florida

[Richard Garlikov](#)

hmwkhelp@scott.net

[Alison I. Griffith](#)

York University

[Ernest R. House](#)

University of Colorado

[Craig B. Howley](#)

Appalachia Educational Laboratory

[Richard M. Jaeger](#)

University of North
Carolina--Greensboro

[Benjamin Levin](#)

University of Manitoba

[Dewayne Matthews](#)

Western Interstate Commission for Higher
Education

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University
