

8. Reporting Policy and Accountability

Alexander, L., & James, H. T. (1987). *The nation's report card: Improving the assessment of student achievement* (Report of the Study Group). Washington, DC: National Academy of Education.

Recommendations of the Study Group on the National Assessment of Educational Progress (NAEP) are summarized in this report. The report is then reviewed by the National Academy of Education. The recommendations are to: (1) maintain NAEP's continuity; (2) assess the core curriculum; (3) focus on transitional grades (4, 8, and 12) and sample out-of-school 17-year-olds, adults and private school students; (4) create an independent Educational Assessment Council, with members to be appointed by the Secretary of Education; (5) provide for add-on and school district assessments; (6) assess and provide for add-on assessment of private school students; and (7) increase federal funding to 20 to 30 million dollars per year (approximately five times the present amount). The Study Group strongly recommends that achievement data be collected on each state and the District of Columbia and that state and local assessments be linked with NAEP. Curriculum areas to be tested include higher order thinking skills; reading, writing and literacy; mathematics, science, and technology; history, geography, and civics; and special topics which are assessed occasionally. Recommendations for measurement include computer assisted testing and scaling. [Authors' abstract]

DeVito, P. J. (1997). The future of the National Assessment of Educational Progress from the states' perspective. In R. Glaser, R. Linn, & G. Bohrnstedt (Eds.), *Assessment in transition: Monitoring the nation's educational progress*. Stanford, CA: National Academy of Education.

Forte Fast, E. (1999, April). *Education indicators and accountability systems: Critical issues in development and reporting*. Invited panelist at the meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Forte Fast, E. (2000, January). *Rationale for selecting educational reporting indicators*. Invited presentation at the Issues in State Accountability: Making Informed Decisions about Accountability Reports workshop, Council of Chief State School Officers, San Antonio, TX.

Forte Fast, E. (2003, April). *Improving reporting and use of data in accountability systems*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Forte Fast, E., & Tucker, C. (2001, April). *Redesign of the student assessment reporting system in Connecticut*. Paper presented at the meeting of the American Educational Research Association, Seattle, Washington.

Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress*. Stanford, CA: National Academy of Education.

Goertz, M. E., Duffy, M. C., & Le Floch, K. C. (2001). *Assessment and accountability systems in the 50 states: 1999-2000* (Consortium for Policy Research in Education Report Series RR-046). University of Pennsylvania, Consortium for Policy Research in Education. Retrieved March 31, 2009, from http://www.cpre.org/Publications/Publications_Accountability.htm

Hamilton, L. S., & Koretz, D. M. (2002). Tests and their use in test-based accountability systems. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 13-49). Santa Monica, CA: RAND. Retrieved March 31, 2009, from https://www.rand.org/pubs/monograph_reports/MR1554/MR1554.ch2.pdf

In this chapter, we provide a short history of large-scale testing and test-based accountability, describe features of the accountability systems and tests that are in place today, and describe several ways in which state testing systems vary. We discuss content and performance standards, which in today's systems typically serve as the means for communicating a common set of goals. Following that, we present information on the features of tests, discuss several other issues related to large-scale testing, including methods of reporting, procedures for setting performance targets, and test-based rewards and sanctions. [Authors' abstract]

Jaeger, R. M. (1992). General issues in reporting of the NAEP trial state assessment results. In R. Glaser & R. L. Linn (Eds.), *Assessing student achievement in the states* (pp. 107-109). Stanford, CA: National Academy of Education.

Jaeger, R. M. (1998). *Reporting the results of the National Assessment of Educational Progress* (NAEP Validity Studies). Washington, DC: American Institutes for Research. Retrieved March 31, 2009, from http://www.air.org/publications/documents/Reporting_NAEP.pdf

This paper explores the ways results of National Assessment of Educational Progress (NAEP) data might be communicated to a variety of audiences, each with differing needs for information, interest in its findings, and sophistication in interpreting the results. The paper describes market-basket reporting as a feasible alternative to traditional NAEP reporting. Such reports would include samples of items and exercises with their scoring rubrics. The second section of the paper makes the case that in order to up-hold the strict standards of data quality, NAEP reports must format and display results to make them more accessible while discouraging readers from drawing overly broad interpretations of the data. A final section describes a detailed program of research on reporting and dissemination of NAEP findings based on these dimensions: (1) the research questions to be asked; (2) the audiences to whom the questions should be addressed; and (3) the strategies through which the questions should be pursued.

The paper suggests that the highest priority be given to research on reporting through public media, followed by making NAEP reporting more understandable and useful to school curriculum and instruction personnel, reporting to the public, and further research with state education personnel. [Author's abstract]

Jaeger, R. M. (2003). *NAEP validity studies: Reporting the results of the National Assessment of Educational Progress* (Working Paper 2003-11). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Jaeger, R. M., Gorney, B., Johnson, R., Putnam, S. E., & Williamson, G. (1994). *A consumer report on school report cards*. Kalamazoo, MI: Western Michigan University, the Evaluation Center.

Jaeger, R. M., Gorney, B., Johnson, R., Putnam, S. E., & Williamson, G. (1994). *Designing and developing effective school report cards: A research synthesis*. Kalamazoo, MI: Western Michigan University, the Evaluation Center.

Jaeger, R. M., & Tucker, C. G. (1998). *Analyzing, disaggregating, reporting, and interpreting students' achievement test results: A guide to practice for Title I and beyond*. Washington, DC: CCSSO.

Jennings, J., & Stark, D. (1997). The future of the National Assessment of Educational Progress. In R. Glaser, R. Linn, & G. Bohrnstedt (Eds.), *Assessment in transition: Monitoring the Nation's Educational Progress*. Stanford, CA: National Academy of Education.

Johnson, E., Lazer, S., & O'Sullivan, C. (1997). *NAEP reconfigured: An integrated redesign of the National Assessment of Educational Progress*. Washington, DC: National Center for Education Statistics.

Chapters in this report outline the potential plans for the redesign of the National Assessment of Educational Progress (NAEP). It is argued that any successful redesign must consider the NAEP as a whole. This report reviews overall NAEP designs and discusses the implications that each of the designs has for various functional areas. [Authors' abstract]

Koretz, D. (1995). The quality of information from NAEP: Two examples of work done in collaboration with Leigh Burstein. *Educational Evaluation and Policy Analysis*, 17(3), 280-294.

This article summarizes two research efforts, both focusing on the mathematics assessments of the National Assessment of Educational Progress, that illustrate Leigh Burstein's long-standing concern with the quality of information about the condition of education. The first examined nonresponse to NAEP test items; it found that omit rates were highest for difficult constructed-response items and that African American and Hispanic students had higher omit rates than Whites.

The second study evaluated the validity of the 1992 achievement level descriptions as characterizations of mathematics performance; it found that the descriptions and accompanying exemplar items were misleading. In response to these findings, a variety of recommendations were offered pertaining to test construction, standards setting, routine monitoring (and reporting) of data quality, and standards-based reporting of student performance. [Author's abstract]

Koretz, D. (1996). Using student assessments for educational accountability. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 171-195). Washington, DC: National Research Council.

Landgraf, K. M. (2001). *Using assessments and accountability to raise student achievement* [On-line]. Retrieved March 31, 2009, from <ftp://ftp.ets.org/pub/corp/kurttest.pdf>

Linn, R. L. (1998). *Assessments and accountability* (CSE Technical Report 490). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Teaching.

Uses of tests and assessments as key elements in five waves of educational reform during the past 50 years are reviewed. These waves include the role of tests in tracking and selection emphasized in the 1950s, the use of tests for program accountability in the 1960s, minimum competency testing programs of the 1970s, school and district accountability of the 1980s, and the standards-based accountability systems of the 1990s. Questions regarding the impact, validity, and generalizability of reported gains and the credibility of results in high-stakes accountability uses are discussed. Emphasis is given to three issues of currently popular accountability systems. These are (a) the role of content standards, (b) the dual goals of high performance standards and common standards for all students, and (c) the validity of accountability models. Some suggestions for dealing with the most severe limitations of accountability are provided. [Author's abstract]

Linn, R. L. (2001). Validation of the uses and interpretations of results of state assessment and accountability systems. In J. Tindal & T. Haladyna (Eds.), *Large-scale assessment programs for all students: Development, implementation, and analysis* (pp. 27-48). Mahwah, NJ: Lawrence Erlbaum Associates.

Provides an overview of validity within the context of assessment and accountability systems mandated and developed by states. Following usage in the 1999 Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council of Measurement in Education; hereafter referred to as Test Standards), the term test is used in a broad sense to include any systematic evaluative device or assessment procedure. The author uses the Test Standards as an organizing tool to discuss the types of evidence and logical arguments that those responsible for state assessment and accountability systems should develop to evaluate the

validity of the uses and interpretations that are made of results. The author begins with a brief overview of the concept of validity and the way in which thinking about the validity in the measurement system has evolved over time. This is followed by a discussion of specific uses and interpretations of results of state assessment and accountability systems and the requirements of the Test Standards to evaluate the validity of those uses and interpretations. [Author's abstract]

Linn, R. L. (in press). Accountability system design. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education*, Vol. 104, Part I.

Linn, R. L., & Dunbar, S. B. (1992). Issues in the design and reporting of the National Assessment of Educational Progress. *Journal of Educational Measurement*, 29(2), 177–194.

Several issues related to the design and reporting of NAEP results are discussed within the context of current expectations for NAEP and its historical origins. Procedures for establishing the content and form of assessments, including the process of developing frameworks, and eventually individual assessment items are discussed. The need to maintain a comprehensive assessment reflecting both current practice in schools and the best thinking by subject matter experts is emphasized. Issues in the design and the estimation of subpopulation parameters using conditioning variables are discussed. Finally, continuing misinterpretations of anchor item results are analyzed. [Authors' abstract]

NAEP Validity Studies Panel. (2002). *An agenda for NAEP validity research*. Washington, DC: American Institutes for Research.

National Assessment Governing Board. (1996). *Redesigning the National Assessment of Education Progress, policy statement*. Washington, DC.

National Assessment Governing Board. (2006). *Policy statement on reporting, release, and dissemination of NAEP results*. Retrieved March 31, 2009, from <http://www.nagb.org/policies/PoliciesPDFs/Reporting%20and%20Dissemination/Reporting,%20Release,%20and%20Dissemination%20of%20NAEP%20Results.pdf>

National Center for Education Statistics. (2000, May). *Rewards for NAEP: Proposals and consequences*. Paper prepared for meeting of the Design and Analysis Committee, Washington, DC.

National Education Goals Panel. (1991). *The national education goals report: Building a nation of leaders*. Washington, DC.

Snodgrass, D., & Salzman, J. A. (2002, April). *Creating the Rosetta stone: Deciphering the language of accountability to improve student performance*. Paper presented

at the meeting of the American Educational Research Association, New Orleans, LA.

The objectives of this paper are threefold. First, a model is proposed for unifying massive amounts of conceptual and numerical information flowing from the measures of the accountability movement in Ohio and the materials that are publicly available to educators. Second, this model is translated into useable forms of information that help teachers drive instructional practices in their classrooms. Third, the efficacy of this model meant to improve state-mandated proficiency scores at the district level is discussed. The paper attempts to integrate numerical, pictorial, graphical and narrative information about the Ohio Proficiency Tests in a way that provides the reader with a rudimentary model of an educational Rosetta Stone. This tool helps educators decipher the contents of the Ohio proficiency tests at a level complex enough so educators can identify basic and fundamental instructional needs of their students. Six appendixes contain sample items from the Ohio proficiency tests. [Authors' abstract]

Twing, J. S. (2008, March). *Score reporting, off-the-shelf assessments and NCLB: Truly an unholy trinity*. Paper presented at the meeting of National Council on Measurement in Education, New York, NY.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2007). *Comprehensive evaluation of NAEP: Utility study final report* (Final Report to Congress). [Also Center for Educational Assessment Report No. 624]. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

7. Sample reports

Individual reports

British Columbia Ministry of Education. (2009). *Foundation skills assessment: Individual student results report*. Victoria, BC: Author. Retrieved March 31, 2009, from http://www.bced.gov.bc.ca/assessment/fsa/results_interpret.htm

The FSA Individual Student Results Report sent home from schools are provided in English. To help families and students for whom English is not the language spoken in the home, translated versions of the Individual Student Results Report are available on this site in 14 additional languages. [Author's abstract]

Connecticut State Board of Education. (2009). *Individual student performance reports: Connecticut mastery test, 4th generation*. Hartford, CT: Author. Retrieved March 21, 2009, from <http://www.ctreports.com/>

CTB/McGraw-Hill. (2003). *Interactive reports page* [On-line]. Retrieved March 31, 2009, from http://www.ctb.com/mktg/terranova/tn_reports_noflash.jsp