

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 12, Number 5, May 2007

ISSN 1531-7714

Using NAEP to Confirm State Test Results in the No Child Left Behind Act

Bert D. Stoneberg
Office of the Idaho State Board of Education

The U.S. Department of Education has not yet published an official guidance document for using NAEP achievement level scores to confirm state testing results. A review of the literature, however, identified four principles that inform the valid use of NAEP scores in a confirming analysis. These principles address the appropriate NAEP statistic to use to confirm state testing results, the difference between NAEP and state definitions of “proficiency,” a rationale for avoiding point-by-point comparisons, and a rationale for using trend analysis.

This paper was conceived when the Idaho State Superintendent of Public Instruction forwarded a question from the Chairman of the Idaho Senate Education Committee. The Senator wanted an explanation for the *large discrepancy* in 2005 between the percentages of Idaho students scoring proficient or better in reading as reported by the state assessment and by the National Assessment of Educational Progress (NAEP). His specific reference was to a press release from the Fordham Foundation that labeled Idaho as one of the worst offenders in the “race to the bottom” by lowering standards and making state tests easier (Leischer, 2005).

The Senator could well have pointed to other reports from well-known individuals and prominent organizations that also advanced variations of the *large discrepancy* theme. Their number is legion. They include papers and articles from associates and staff of the Brookings Institution (Ravitch, 2005), the Center for American Progress (Rocha & Brown, 2005), the Education Trust (Hall & Kennedy, 2006), the Hoover Institution (Finn & Ravitch, 2006; Peterson & Hess, 2006), the National Center for Research on Evaluation, Standards, and Student Testing (Linn, Baker & Herman, 2005), the Northwest Regional Educational Laboratory (Greenough, 2005), Policy Analysis for California

Education (Fuller, Gesicki, Kang & Wright, 2006), and the Rand Corporation (McCombs & Carrol, 2005). This list is illustrative, not exhaustive. Unfortunately the authors of these reports were either unaware of or elected to ignore published information related to the valid use of NAEP achievement level scores to confirm state test results.

The U.S. Department of Education is responsible to provide guidance about how NAEP scores are intended to be interpreted and used.

The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education have collaborated to establish and publish professional standards for educational and psychological testing (Joint Committee, 1999). Two of the professional standards relating to the valid use of test data:

- Standard 1.2. The test developer should set forth clearly how test scores are intended to be interpreted and used.
- Standard 1.4. If a test is used in a way that has not been validated, it is incumbent on the user

to justify the new use, collecting new evidence if necessary.

Congress mandated NAEP and assigned the role of test developer to the U.S. Department of Education (ED). This being the case, Standard 1.2 places responsibility upon ED to specify the appropriate interpretation and use of NAEP achievement level scores. The lead ED groups responsible for NAEP are the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES). The Office of Elementary and Secondary Education oversees *No Child Left Behind Act* (NCLB) program issues and implementation.

NAEP has been able to report state-level results since 1990. When the *No Child Left Behind Act* became law in 2001, it suggested a new use for NAEP achievement level scores. Indeed, the law indicated that the Secretary of Education may use NAEP scores, in conjunction with other measures, to confirm a state's adequate yearly progress (AYP). ED has not yet published a set of "how to" guidelines for the valid use of NAEP achievement level scores to confirm state test results. This lack of an official guidelines document from the developer, however, does not constitute license to use NAEP scores haphazardly or without caution. There is sufficient information about the topic in publications from ED and from external sources that one can identify a few basic principles or "ground rules."

NAEP releases state-level summary data to the public in both paper and electronic formats. Credentialed educational researchers can obtain access to the raw data. Standard 1.4 leaves the burden to justify any new use of NAEP data that has not been previously validated upon the user.

Principle 1: The percentage at or above *Basic* is the appropriate NAEP statistic for confirming state AYP results.

In 2004, the NAEP Validity Studies Panel published a finding that the percent at or above *Basic* is the most appropriate NAEP statistic to use when confirming state AYP results. The Panel was established by NCES via a contract with the American Institutes for Research (AIR) to provide technical reviews of NAEP but its publications represent the views of the authors, and not necessarily the views of AIR or NCES. From the NAEP Validity Studies Panel's report:

Adequate yearly progress is already defined within the Act based on the percentage of scores exceeding the basic proficiency level. The basic proficiency level corresponds roughly to the

percentage below basic on the NAEP scale. Therefore, of the various statistics that might be used for measuring a gap on the NAEP scale—proportion at or above the basic, proficient, or advanced achievement level, or mean standardized score—the proportion at or above the basic achievement level will both have the greatest correlation with the adequate yearly progress statistic and also be the most directly comparable. Since gaps and AYP measure different performance objectives (equality vs. absolute improvement), it follows that using the same basic statistic to measure each would simplify both interpretation and the presentation of results (Mosquin & Chromy, 2004).

Narratives, tables and charts in NAEP reports that NCES prepared for the 2003 and earlier state-level assessments focused exclusively on the percent of students at or above *Proficient*. In reports for NAEP 2005 some of the narratives, tables and charts prominently displayed the percent of students at or above *Basic* for the first time. For example, the graphic showing the percentage of students within each reading achievement level by state in the *Nation's Report Card* for NAEP 2003 focused on at or above *Proficient* (Donahue, Daane, & Grigg, 2003). The corresponding state level graphic in the *Nation's Report Card* for the 2005 assessment drew attention to at or above *Basic* (Perie, Grigg, & Donahue, 2005). This noteworthy change in NCES's reporting practices for NAEP seems to concur with the Panel's findings.

Figure 1 illustrates "side-by-side" the visual impact of shifting attention from at or above *Proficient* to at or above *Basic* when graphing the percentage of students by achievement level. It displays both NAEP statistics for Idaho fourth-grade students within each mathematics achievement level from the assessments in 1992, 2000, 2003 and 2005. Student performance is identical in both graphs, only the focus has changed. These data and graphics for Idaho were generated by the NAEP Data Explorer, which is an online, interactive tool that the National Center for Education Statistics makes available for public use to mine the NAEP database for state and national results (U.S. Department of Education, 2005).

In Figure 1 the fourth-grade mathematics performance trend over time for both NAEP statistics is positive. The outcome of an analysis using NAEP's percent at or above *Proficient* to confirm state AYP results will correlate with outcome of an analysis based on NAEP's percent at or above *Basic*. Nonetheless, as Mosquin and Chromy (2004) stated, "the proportion at or above the basic achievement level will both have the greatest correlation with the adequate yearly progress statistic and also be the most directly comparable."

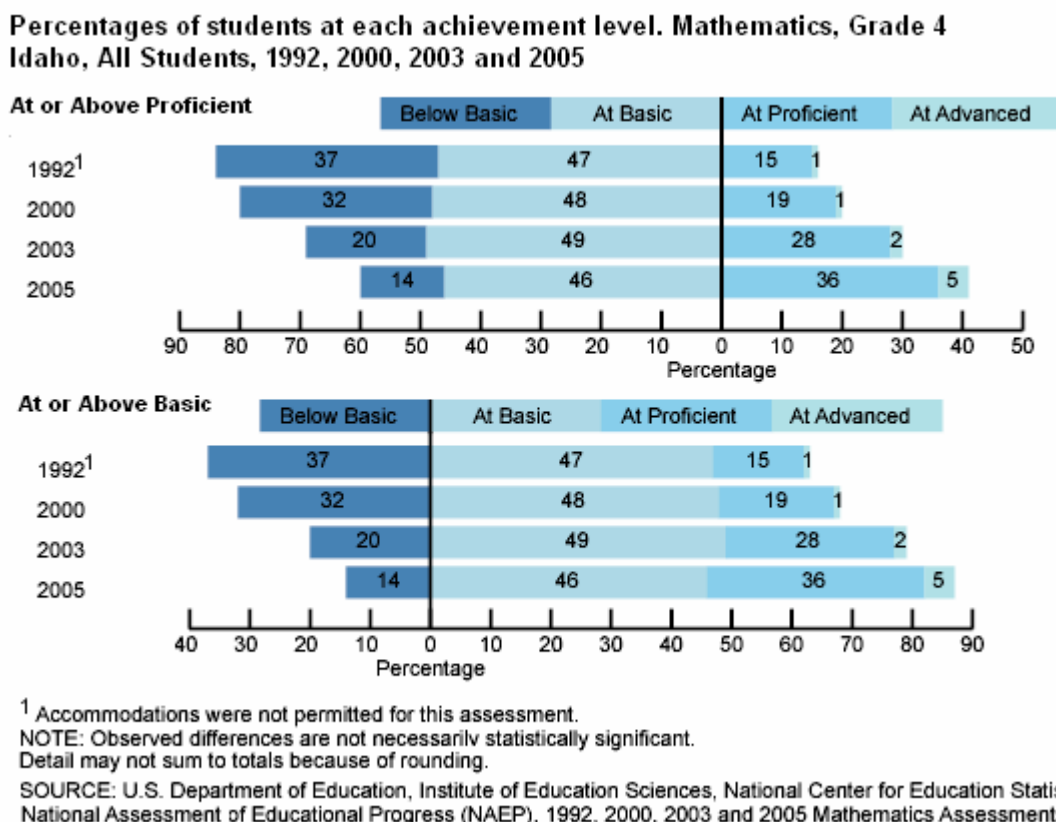


Figure 1. An illustration of the visual impact of shifting focus of attention from at or above *Proficient* to at or above *Basic* when graphing the percentage of Idaho fourth-grade students within each mathematics achievement level in 1992, 2000, 2003 and 2005.

Principle 2: The NAEP definition of *Proficient* is not synonymous with “proficiency” in a subject.

U.S. Department of Education publications such as the achievement level booklets and the assessment frameworks provide and explain the definition of NAEP *Proficient*.

Achievement Level Booklets. NAGB has published a series of booklets to inform the general public about the use and interpretation of NAEP achievement level scores. Each booklet discusses a separate subject for which NAEP achievement levels have been established. These include reading, mathematics, science, writing, civics, U.S. history, and geography. The following text is from the reading booklet section entitled *How Should Achievement Levels Be Interpreted*, but identical language appears in all seven booklets:

Achievement levels define performance, not students. Notice that there is no mention of “at grade level” performance in these achievement goals. In particular, it is important to understand

clearly that the Proficient achievement level does not refer to “at grade” performance. Nor is performance at the Proficient level synonymous with “proficiency” in the subject. That is, students who may be considered proficient in a subject, given the common usage of the term, might not satisfy the requirements for performance at the NAEP achievement level. Further, Basic achievement is more than minimal competency. Basic achievement is less than mastery but more than the lowest level of performance on NAEP. Finally, even the best students you know may not meet the requirements for Advanced performance on NAEP. (Loomis & Bourque, 2001b).

The NAEP definition of *Proficient*, as stipulated, is technical and is not synonymous with grade-level proficiency in a subject. By contrast, NCLB requires the states to focus on grade-level performance. “We remain committed to ensuring that all students can read and do math at grade level or better by 2014. This is the basic purpose and mission of the No Child Left Behind Act” (U.S. Department of Education, 2007).

NAEP *Proficient* is not synonymous with proficiency in the subject given the common usage of the term. By

contrast, under NCLB state assessment programs must measure and report proficiency in the subject. One criterion a state must demonstrate to pass a federal Peer Review of its testing program is, “The State’s academic achievement standards fully reflect its academic content standards for each required grade and describe what content-based expectations each achievement level represents. The ‘proficient’ achievement level represents attainment of grade-level expectations for that academic content area” (U.S. Department of Education, 2004).

Assessment Frameworks. NAGB has also published an assessment framework for each subject that includes an extended description of the NAEP achievement levels. From the framework used for NAEP 2005, the text below describes the *Basic* achievement level for fourth grade reading:

Fourth-grade students performing at the Basic level should demonstrate an understanding of the overall meaning of what they read. When reading text appropriate for fourth graders, they should be able to make relatively obvious connections between the text and their own experiences and extend the ideas in the text by making simple inferences.

For example, when reading literary text, they should be able to tell what the story is generally about—providing details to support their understanding—and be able to connect aspects of the stories to their own experiences.

When reading informational text, Basic-level fourth graders should be able to tell what the selection is generally about or identify the purpose for reading it, provide details to support their understanding, and connect ideas from the text to their

background knowledge and experiences. (Reading Framework, 2004).

The pre-publication edition of the framework for the 2009 national assessment of reading leaves no doubt that NAEP *Proficient* is different from expected grade-level performance. “Proficient readers will have sizable meaning vocabularies, including knowledge of many words and terms above grade level” (American Institutes for Research, 2007).

Table 1 represents the author’s attempt to understand the NAEP achievement levels by “matching” NAEP’s descriptive language to a range of letter grades that one might see on the report cards of students performing at each NAEP achievement level. The letter grades are based upon the author’s thirty years of experience in the public schools of Washington, Oregon and Idaho, and upon a hazy, general awareness about how students seem to be distributed across the achievement levels and letter grades. The reader, of course, is free to estimate different grade ranges for the NAEP achievement levels based on their own experience.

Principle 3: Confirmation of state testing results should not be conducted on a point-by-point basis.

NAGB convened an Ad Hoc committee to study how NAEP might best be used to confirm state test AYP results. The committee concluded, “Informed judgment’

Table 1. U.S. Department of Education English language descriptors for each NAEP achievement level in the reading achievement level reports and reading frameworks, and an estimated range of “letter grades” describing each NAEP achievement level.

NAEP Achievement Level	NAEP English Language Descriptor	Range of Grades
<i>Advanced</i>		TAG A+
<i>Proficient</i>	Some of the best students you know Many words and terms above grade level Mastery	A B+
<i>Basic</i>	Proficiency in subject (common meaning) Overall understanding of grade-appropriate text More than minimal competency	B C-
<i>Below Basic</i>	Minimally competent	D+ F

and a ‘reasonable person’ standard should be applied in using National Assessment data as confirmatory evidence for state results. Confirmation should not be conducted on a ‘point by point’ basis or construed as a strict ‘validation’ of the state’s test results” (Ad Hoc Committee, 2002).

Point-by-Point Confirmation: Within State. In Idaho, the percentage of students at or above *Basic* on NAEP is often in the neighborhood of the percentage of students at or above proficient on the state test. In 2005, as illustrated in Figure 2, 76 percent of the eighth graders were at or above *Basic* for NAEP reading while 82 percent were at or above proficient on the state test. For mathematics, 76 percent were at or above *Basic* on NAEP while 69 percent were at or above proficient on the state test. A point-by-point comparison of these state and NAEP scores could mislead one to seemingly “obvious” but nonetheless dubious claims. Even when using the most directly comparable state AYP and

NAEP statistics, as in Figure 2, there is no support in the observed point differences to claim that Idaho’s reading test was less rigorous than the NAEP reading test, or that Idaho’s mathematics test was more rigorous than the NAEP mathematics test.

As a part of any confirming analysis, reasonable persons exercising informed judgment will explore and report the major differences between the state test and NAEP. Such differences include, but certainly are not limited to those listed in Table 2. The Ad Hoc Committee stressed, "Potential differences may be minimal or great in number and in size. They cannot reasonably be expected to operate in all states in equal fashion. The greater the differences between a state test and NAEP, the greater the complexity in using NAEP as confirmatory evidence for the state’s test results, and the greater the cautions in interpretation that should accompany the confirmatory evidence."

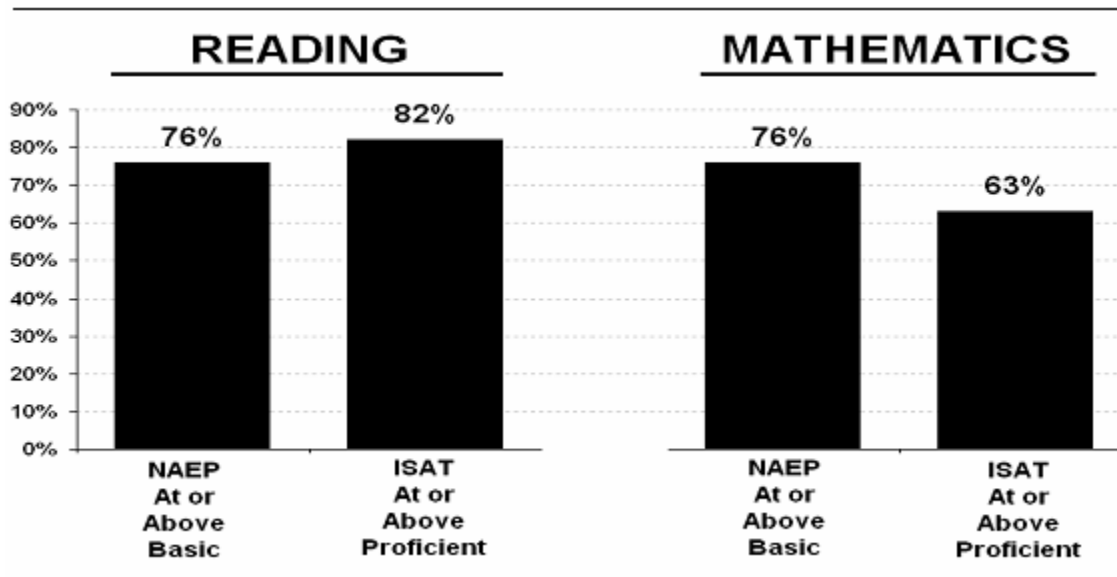


Figure 2. The percentage of Idaho eighth graders at or above *Basic* on NAEP and at or above proficient on the ISAT (Idaho Standards Achievement Test) for reading and mathematics in 2005

The use of NAEP achievement levels to confirm a nationally standardized test (SAT, ACT, etc.) has the same limitations as using NAEP to confirm a state test. Simons and Mwalimu (2000) conducted a study for NAGB that assembled four focus groups to collect public comment regarding the criteria for NAEP achievement levels. The homogenous groups consisted of (1) governors’ and states’ legislative staffs, (2) state

assessment personnel, (3) public and private school teachers and administrators, and (4) parents, business leaders and education policymakers. All four discussion groups agreed that the NAEP achievement levels cannot be compared with results from other standardized national assessments citing differences in sampling methodologies and student motivations for taking the tests.

Table 2. Elements, characteristics and environment of testing programs in which a state test may differ sufficiently enough from NAEP to require cautions in the interpretation of confirmatory evidence.

<ul style="list-style-type: none"> • assessment design (one form, multiple forms, alternate forms) • content coverage in the state standards (and the NAEP framework) • definition of reporting subgroups • frequency of administration (annual, fall and spring, etc.) • history of the test (new test vs. long-term well-known test, etc.) • information or practices unique to a state • method of presentation (computer, pencil-and-paper, etc.) • mix of item formats (multiple choice, constructed response, etc.) • number of items per assessment • parent notification or permission requirements • principal or teacher expectations for students taking the tests 	<ul style="list-style-type: none"> • reporting metric (raw score, scale score, achievement levels, etc.) • sampling procedure and rules for excluding students from testing • sequencing of content in the subjects • standard-setting approaches • state demographics, including changes over time • state policies or laws regulating the state assessment • student motivation in taking the state test versus taking NAEP • test difficulty and range of item difficulties • tools student may use during the test (dictionary, calculator, etc.) • type of assessment (criterion related, norm referenced, portfolio, etc.) • whether the test is timed
--	---

Point-by-Point Confirmation: Cross-State. Cross-state point-by-point confirming analyses also mislead one to seemingly “obvious” claims that are, as a matter of fact, without merit. The Fordham Foundation’s study (Leischer, 2005), for example, focused on the point-by-point difference between the change from 2003 to 2005 in the percentage of students scoring proficient on the state test and the corresponding change in percentage of students scoring *Proficient* on NAEP. When the difference on a state test was higher than on NAEP, it was attributed solely to state educators and politicians “blurring the truth to make themselves look better.” The study was conspicuously silent about differences between the state tests and how each state’s test differs from NAEP. There was no mention of the potential impact that these differences might have on the interpretation of the points-by-points comparisons.

Most of the *large discrepancy* reports cited above implemented one variation or another of point-by-point methodology. According to the fourth grade reading scores in Table 3 (data selected from Hall & Kennedy, 2006), the point difference between the percentage at or above proficient on the state test and the percentage at or above *Proficient* on NAEP was smaller in Delaware (51) than in Idaho (54). In the absence of complete disclosure about differences between the two state reading tests, it might appear to some that Delaware’s test was more rigorous than Idaho’s.

The Fordham Foundation study also rank ordered the states on the differences between the state and NAEP percentages to identify the “worst offenders.” As a consequence of the sampling procedure, however, there is sufficient quantifiable uncertainty in NAEP scores that they should not be used to rank order the states (Stoneberg, 2005). Indeed, the 95 percent confidence intervals for the NAEP scores (i.e., score \pm 1.96 times the standard error) overlapped for the five states listed in Table 3. So while Delaware’s 34 percent at or above *Proficient* on NAEP may appear to be higher than Oregon’s 29 percent, there was no statistically significant difference between the two percentages. The uncertainty in the NAEP scores alone could account for virtually all the variation in the observed differences between the state and NAEP percentages.

Furthermore, the number and size of potential differences between the state tests make it next to impossible to demonstrate whether the 87 percent proficient or higher on South Dakota’s reading test in Table 3 describes essentially the same knowledge and skills attainment as the 87 percent proficient or higher on Idaho’s test. With all of the uncertainty within and between the 50 state tests and NAEP, it makes little sense to compare or rank the rigor of fifty state tests based on the difference between NAEP scores and the corresponding state AYP scores. Such a comparison seems more likely to compound uncertainty and error than to reduce them.

Table 3. Ranking the rigor of five state reading tests by comparing the difference between the state percentage at or above proficient and the NAEP percentage at or above *Proficient*, Grade 4, 2005.

	State % Proficient (1)	NAEP % <i>Proficient</i> (2)	Difference (1)-(2)	Rank (1)-(2)
Delaware	85	34(1.2)	51	1
Idaho	87	33(1.4)	54	4
North Carolina	83	29(1.4)	54	4
Oregon	81	29(1.5)	52	2
South Dakota	87	33(1.3)	54	4

Note: Standard errors for NAEP are enclosed in parentheses. Scores are taken from a report by The Education Trust (Hall & Kennedy, 2006).

Principle 4: Confirmation should not be construed as a strict validation of the state’s test results.

The use of achievement levels is a developing process and is subject to various interpretations. When NAEP sets achievement levels it pays great attention to detail and technical precision. Well-qualified people from across the nation take part in a multi-day process that includes training to prepare them to make realistic judgments about student performance on NAEP items. Sophisticated psychometric methods guide the process. “Extensive analyses are conducted to determine whether panelists seemed to be making logical, informed judgments and whether similar panelists would make similar judgments. Yet, there is no way of knowing that the standards are ‘right’ because there is no true standard against which to evaluate the panelists’ judgments” (Loomis & Bourque, 2001b).

Congress has mandated external evaluations of NAEP, the most recent of which was conducted by the National Academy of Sciences (NAS). The NAS panel’s conclusions on the topic of reasonable and useful performance standards for NAEP included (Pellegrino, Jones & Mitchell, 1998):

- Standard setting rests on informed judgment, but the complexity of NAEP’s current achievement-level-setting procedures can create the misleading impression that level setting is a highly objective process, rather than a judgmental one.

- NAEP’s current achievement-level-setting procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; rater’s judgments of different item types are internally inconsistent; appropriate validity evidence for the cutscores is lacking; and the process has produced unreasonable results.

The NAS panel’s recommendations included:

- NAEP’s current achievement levels should continue to be used on a developmental basis only. If achievement-level results continue to be reported for future administrations of assessments in which achievement levels have already been set, the reports should strongly and clearly emphasize that the achievement levels are still under development, and should be interpreted and used with caution.
- Reports should focus on the change, from one administration of the assessment to the next, in the percentages of students in each of the categories determined by the existing achievement-level cutscores (below basic, basic, proficient, and advanced), rather than focusing on the percentages in each category in a single year.

It is noteworthy that even though the NAS panel was generally critical of NAEP’s achievement levels, it did recommend their use for drawing attention to changes in student performance over time. This recommendation is entirely consistent with NAEP’s

mission, which is to measure student achievement and to report change in performance over time.

Given the current status of NAEP achievement levels and the stated purpose of the national assessment, it may be that graphing trend lines showing state percent at or above proficient and NAEP percent at or above *Basic* together offers the most defensible method for confirming state AYP results. If the trend lines are moving in the same direction, it may be said that NAEP confirms the state results. This is not “strict validation,” but it works.

After thoroughly reviewing what NAEP had learned over a decade about what works and does not work for large-scale assessment programs, Loomis and Bourque (2001a) concluded, “The standard-setting movement is marching ahead. At this point, the policy demand to set standards may be ahead of the technology resources to set them.”

Discussion

Inquiries about the *large discrepancy* issue have come not only from the Senator but from the Idaho Statehouse, from the State Department of Education, from district and school administrators, and from classroom teachers. The answer to all -- from senator to teacher -- has been the same. The authors of the cited reports were apparently either unaware of or chose to ignore available guidance regarding the valid use of NAEP scores to confirm state testing results. Had they exercised due caution in their analyses of the 2005 data (i.e., had they paid attention at least to the principles or “ground rules” identified in this paper) their findings undoubtedly would have been different, if not opposite.

Nothing in this paper should be construed as a criticism of the National Assessment of Educational Progress or of any of the state testing programs. The criticism is directed at the failure of researchers to make a valid use of NAEP achievement level scores to confirm state AYP results. Consider two examples illustrating how this failure has actually harmed the American education community.

First, the U.S. Chamber of Commerce distributed to its 3,000,000 plus members a report from one of its affiliates entitled *Leaders and laggards: A state-by-state report card on educational effectiveness* (Institute for a Competitive Workforce, 2007). One “letter grade” on the report card was for “Truth in Advertising about Student Proficiency.” The Institute for a Competitive Workforce did not itself calculate the states’ grades, but relied on a research report issued by the Hoover Institution (Peterson & Hess, 2006). The methodology consisted of ranking the states on the point-by-point differences

between their percentage of students at or above NAEP *Proficient* and their percentage at or above state proficient. In short, the U.S. Chamber of Commerce and its affiliate awarded Idaho, for example, a “D” for truth in advertising about student proficiency based on results from a questionable “research report,” unduly undermining community trust in the public schools.

Second, in January, 2007, an *Education Week* article indicated that Congress was considering bills to amend the *No Child Left Behind Act*. One bill encourages states to benchmark their own standards and tests to NAEP. Another would provide incentives for states to adopt voluntary “American education content standards” in mathematics and science that NAGB would develop. To what did the article attribute this political activity? “Studies over the past year have found that, in many states, a far higher percentage of students score at the proficient level on state tests than on NAEP. That’s led to concerns that states’ standards and tests may not be stringent enough, and that pressure to meet achievement targets under the NCLB law may be having the perverse incentive of encouraging states to lower their standards” (Olson, 2007). This is a clear example showing how questionable “research” harms schools through misinforming the development of policies and laws that regulate their daily operations.

NAEP releases state-level results for reading, mathematics, science or writing from the national assessment every two years. It is reasonable to expect that researchers who use NAEP achievement level scores adhere to the published guidelines for their interpretation and use. Researchers (and publishers, editors, columnists, reporters and public information officers who transmit their findings to the world) are encouraged to become aware of and to give due consideration to the principles and “ground rules” for using NAEP achievement level scores available in the literature. Then discussion of findings from future rounds of confirming analyses can focus on genuine educational issues rather than on the quality and usefulness of the findings or lack thereof.

References

- Ad Hoc Committee on Confirming Test Results. (2002). *Using the National Assessment of Educational Progress to confirm state test results*. Washington, D.C.: National Assessment Governing Board. Retrieved December 4, 2006, from http://www.nagb.org/pubs/color_document.pdf
- American Institutes for Research. (2007). *Reading Framework for the 2009 National Assessment of Educational Progress (Pre-Publication Edition)*. Washington, D.C.: National Assessment Governing

- Board. Retrieved March 27, 2007 from http://www.nagb.org/frameworks/reading_fw_03_07_prepub_edition.doc
- Donahue, P., Daane, M., and Grigg, W. (2003). *The Nation's Report Card: Reading highlights 2003*. (NCES 2004-452). Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved December 4, 2006, from <http://nces.ed.gov/nationsreportcard/pdf/main2003/2004452.pdf>
- Finn, C.E., and Ravitch, D. (2006, February 27). Basic instinct. *The Wall Street Journal*, p.A14.
- Fuller, B., Gesicki, K., Kang, E., and Wright, J. (2006). *Is the No Child Left Behind Act working? The reliability of how states track achievement*. Berkeley, CA: Policy Analysis for California Education. Retrieved December 4, 2006, from http://pace.berkeley.edu/NCLB/WP06-01_Web.pdf
- Greenough, R. (2005). Region at a Glance: What the test scores do—and don't—reveal. *Northwest Education*, 11(2). Retrieved December 4, 2006, from <http://www.nwrel.org/nwedu/11-02/region/>
- Hall, D., and Kennedy, S. (2006). *Primary progress, secondary challenge: A state by state look at student achievement patterns*. Washington, D.C.: The Education Trust. Retrieved December 4, 2006, from <http://www2.edtrust.org/NR/rdonlyres/15B22876-20C8-47B8-9AF4-FAB148A225AC/0/PPSCreport.pdf>
- Institute for a Competitive Workforce. (2007). *Leaders and laggards: A state-by-state report card on educational effectiveness*. Washington, D.C.: U.S. Chamber of Commerce. Retrieved March 1, 2007, from <http://www.uschamber.com/icw/reportcard>
- Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological Testing*. Washington, D.C.: American Educational Research Association.
- Leischer, J. (2005, October 19). *Press release: Gains on state reading tests evaporate on 2005 NAEP*. Washington, D.C.: The Thomas B. Fordham Foundation. Retrieved December 4, 2006, from http://www.edexcellence.net/foundation/about/press_release.cfm?id=19
- Linn, R.L., Baker, E.L., and Herman, J.L. (2005). From the directors: Chickens come home to roost. *CRESST Line (Newsletter of the National Center for Research on Evaluation, Standards, and Student Testing)*, Fall. Retrieved December 4, 2006, from http://www.cse.ucla.edu/products/newsletters_set.htm
- Loomis, S.C., and Bourque, M.L. (2001a). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In Cizek, G.J. (Ed). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Loomis, S.C., and Bourque, M.L. (Eds.) (2001b). *National Assessment of Educational Progress achievement levels 1992-1998 for reading*. Washington, D.C.: National Assessment Governing Board. Retrieved December 4, 2006, from <http://www.nagb.org/pubs/readingbook.pdf>
- McCombs, J.S., and Carroll, S.J. (2005). Ultimate test: Who is accountable for education if everybody fails? *Rand Review*, 29(1). Retrieved December 4, 2006, from <http://www.rand.org/publications/randreview/issues/spring2005/ulttest.html>
- Mosquin, P., and Chromy J. (2004). *Federal sample sizes for confirmation of state tests in the No Child Left Behind Act*. Washington, D.C.: American Institutes for Research, NAEP Validity Studies Panel. Retrieved December 4, 2006, from http://www.air.org/publications/documents/MosquinChromy_AIR1.pdf
- Olson, L. (2007, January 9). New bills would prod states to take national view on standards. *Education Week*, Volume 26, Issue Web Only. Retrieved January 16, 2007, from http://www.edweek.org//ew/articles/2007/01/09/18stand_web.h26.html?print=1
- Pellegrino, J.W., Jones, L.R., and Mitchell, K.J. (Eds.). (1998). *Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press. Retrieved December 4, 2006, from <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED446096>
- Perie, M., Grigg, W., and Donahue, P. (2005). *The Nation's Report Card: Reading 2005*. (NCES 2006-451). Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved December 4, 2006, from <http://nces.ed.gov/nationsreportcard/pdf/main2005/2006451.pdf>
- Peterson, P.E., and Hess, F.M. (2006). Keeping an eye on state standards: A race to the bottom? *Education Next*, 2006(3). Retrieved March 1, 2007, from http://media.hoover.org/documents/ednext20063_28.pdf
- Ravitch, D. (2005, November 7). Every state left behind. *The New York Times*, p. A23. Retrieved December 4,

- 2006 from <http://www.brookings.edu/views/oped/ravitch/20051107.htm>
- Reading framework for the 2005 National Assessment of Educational Progress. (2004). Washington, D.C.: National Assessment Governing Board. Retrieved December 4, 2006, from http://www.nagb.org/pubs/r_framework_05/7615_07-ReadingFramework.pdf
- Rocha, E., and Brown, C.G. (2005). *The case for national standards, accountability, and fiscal equity*. Washington, D.C.: Center for American Progress. Retrieved December 4, 2006, from <http://www.americanprogress.org/issues/kfiles/bl164323.html>
- Simmons, C., and Mwalimu, M. (2000). What NAEP Publics Have To Say. In Bourque, M.L. and Byrd, S. (Eds.) *Student performance standards on the National Assessment of Educational Progress: Affirmations and improvements*. Washington, DC: National Assessment Governing Board. Retrieved December 4, 2006, from <http://www.nagb.org/pubs/studentperfstandard.pdf>
- Stoneberg, B.D. (2005). Please don't use NAEP scores to rank order the 50 states. *Practical Assessment, Research, & Evaluation*, 10(9). Retrieved on December 4, 2006, from <http://pareonline.net/getvn.asp?v=10&n=9>
- U.S. Department of Education. (2004). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*. Washington, D.C.: Author. Retrieved March 6, 2007, from <http://www.ed.gov/policy/elsec/guid/saaprguidance.doc>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2005). *NAEP Data Explorer*. Retrieved on December 4, 2006, from <http://nces.ed.gov/nationsreportcard/nde/>
- U.S. Department of Education. (2007). *Building on Results: A Blueprint for Strengthening the No Child Left Behind Act*. Washington, D.C.: Author. Retrieved March 6, 2007, from <http://www.ed.gov/policy/elsec/leg/nclb/buildingonresults.pdf>

Citation

Stoneberg, Bert D. (2007). Using NAEP to Confirm State Test Results in the No Child Left Behind Act. *Practical Assessment Research & Evaluation*, 12(5). Available online: <http://pareonline.net/getvn.asp?v=12&n=5>

Author

Bert D. Stoneberg is the NAEP Program Manager in the Office of the State Board of Education, Idaho. Correspondence regarding this article should be sent to him at the Office of the State Board of Education, 650 West State Street, P.O. Box 83720, Boise, ID 83720-0037 or Bert.Stoneberg@osbe.idaho.gov.