

Council of Chief State School Officers
37th Annual National Conference on Large-Scale Assessment
Nashville, Tennessee
June 19, 2007

An Explanation for the Large Differences between State and NAEP “Proficiency” Scores Reported for Reading in 2005

Bert D. Stoneberg
NAEP State Coordinator
Idaho State Board of Education

Abstract

The *No Child Left Behind Act* (NCLB) permits the Secretary of Education to use NAEP achievement level scores, in concert with other data, to confirm state testing results. The U.S. Department of Education has not yet published a guidance document describing how NAEP might be used appropriately. A review of the literature from the Department and other sources, however, identified information regarding definitions and methodology sufficient to inform interested researchers about the valid use of NAEP for confirming state test results. The NCLB definition of proficiency that states must implement in their testing programs centers on grade-level expectations, but the NAEP definition requires the mastery of challenging content that may include some above-grade-level content. In conducting a valid analysis (1) NAEP *Basic* should be used, not NAEP *Proficient*, (2) point-by-point comparisons should be avoided, (3) differences between NAEP and the state testing program should be documented, and (4) the investigation of state vs. NAEP trends is useful. The researchers and authors of the reports that claimed to find large differences between state and NAEP “proficiency” scores from 2005 were either unaware of or chose to ignore guidance that was available in the literature. The explanation resides in faulty research methodology that ended up with highly questionable, if not out-and-out wrong findings. (22 references, 1 table, 3 figures, and 1 Power Point handout).

Council of Chief State School Officers
37th Annual National Conference on Large-Scale Assessment
Nashville, Tennessee
June 19, 2007

An Explanation for the Large Differences between State and NAEP “Proficiency” Scores Reported for Reading in 2005

Bert D. Stoneberg
NAEP State Coordinator
Idaho State Board of Education

Introduction

The search for this explanation was triggered 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 differences in 2005 between the percentage of Idaho students scoring proficient or better 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 Fordham research assumed that “state proficient” and “NAEP *Proficient*” shared a common definition. The method consisted of calculating growth scores (i.e., change in the percentages of proficient students from 2003 to 2005 on the state tests and on NAEP) and rank ordering the states according to the point-by-point comparisons between the state and NAEP growth scores. Fordham reported large differences and attributed any differences favoring a state’s test to state efforts to dumb down their testing programs to make the state look good to the public.

The Fordham Foundation was not the only voice proclaiming large differences between state and NAEP proficiency scores. Other well-known “think tanks” and “researchers” advancing the assertion included 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); the Policy Analysis for California Education (Fuller, Gesicki, Kang & Wright, 2006); the Rand Corporation (McCombs & Carrol, 2005); and the U.S. Chamber of Commerce (Institute for a Competitive Workforce, 2007). This is only an abbreviated list.

Review of Literature

A review of the literature identified three issues related to the valid use of NAEP scores for confirming state test results that the authors of the cited reports were either unaware of or elected to ignore. First, the U.S. Department of Education is responsible for stipulating how NAEP scores are to be used. Second, the NCLB and NAEP programs have employed different definitions for the term “proficiency.” Third, guidelines have been published that espouse dos and don’ts concerning the valid use of NAEP for confirming state test results.

Guidance for Using NAEP Scores

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 standards dealing with the valid use of test data are

- 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 created NAEP and gave it to the U.S. Department of Education. Thus under Standard 1.2, the Department is responsible to specify the appropriate interpretation and use of NAEP scores. The groups in the Department responsible for implementing NAEP are the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB). The Office of Elementary and Secondary Education in the U.S. Department of Education oversees NCLB.

The Department has not published a set of “how to” guidelines for the valid use of NAEP to confirm state test results. This lack of an official guidance document from the test developer, however, does not constitute license to use NAEP achievement level scores without due caution. There is guidance enough in Department publications and from relevant external sources to conduct a valid confirming analysis. Nonetheless, Standard 1.4 permits novel uses of NAEP data but places the burden upon the educational researcher to justify any use NAEP data that has not been previously validated.

Incompatible Definitions of “Proficiency”

The NCLB Definition. The definition of “proficiency” that the U.S. Department of Education uses for NCLB purposes focuses attention on the attainment of grade-level skills in the subject area. A state must implement this definition in its testing program in order to qualify for federal Title I money.

Standards under this paragraph shall “(II) describe two levels of high achievement (proficient and advanced) that determine how well children are mastering the material in the State academic content standards” (*No Child Left Behind Act*, 2001).

NCLB requires that a state attend to proficiency in a subject. One criterion a state must meet for a Peer Review of its testing program reads, “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).

“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). Indeed, it may be said that the basic purpose and mission of the No Child Left Behind Act is to ensure that all students can read and do math at grade level (i.e., proficient) or better (i.e., advanced) by 2014.

The NAEP Definition. NAGB has published a series of booklets to inform the public about the use and interpretation of NAEP scores. The following text is from the reading booklet, but the language is also in the booklets for writing, mathematics, science, U.S. history, geography, and civics:

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, 2001).

NAGB has included descriptions of the achievement levels in the assessment frameworks. For example, text from the NAEP 2005 framework 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 2009 reading framework leaves no doubt that NAEP *Proficient* is different from the expected grade-level performance that NCLB mandates. “Proficient readers will have sizable meaning vocabularies, including knowledge of many words and terms above grade level” (American Institutes for Research, 2007).

These NAGB publications make it clear that NAEP *Proficient* is not restricted to grade-level performance, nor is it synonymous with proficiency in a subject.

Table 1 displays a non-empirical attempt to better define the achievement levels by linking a “letter grade” with the language that NAEP associates with each achievement level. The grades are based on the presenter’s general perception about how student performance is distributed across the NAEP achievement levels and on his personal experience in the public schools of Washington, Oregon and Idaho. Feel free to adjust them to reflect your own experience.

Methodological Guidelines

Use NAEP *Basic*. The NAEP Validity Studies Panel has determined that the percent at or above *Basic* is the appropriate NAEP statistic to use when confirming state AYP results. The Panel was established when NCES contracted with the American Institutes for Research (AIR) to provide technical reviews of NAEP, but its publications represent the views of the research authors, and not necessarily the views of NCES or AIR. The panel has noted:

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

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	Letter 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

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 reports that NAEP released for the 2003 and earlier state-level assessments focused primarily on the percent of students at or above *Proficient*. In reports that NCES issued for NAEP 2005 some of the narratives, tables and charts prominently displayed the percent of students at or above *Basic* for the first time. Figure 1 illustrates the change of focus on trend charts from the fourth grade mathematics snapshot reports for Idaho between 2003 and 2005. This notable change in NCES’s reporting practices for NAEP seems to concur with findings of the NAEP Validity Studies Panel.

Percentages at NAEP Achievement Levels from NCES State Snapshot Reports Idaho, Grade 4, Mathematics

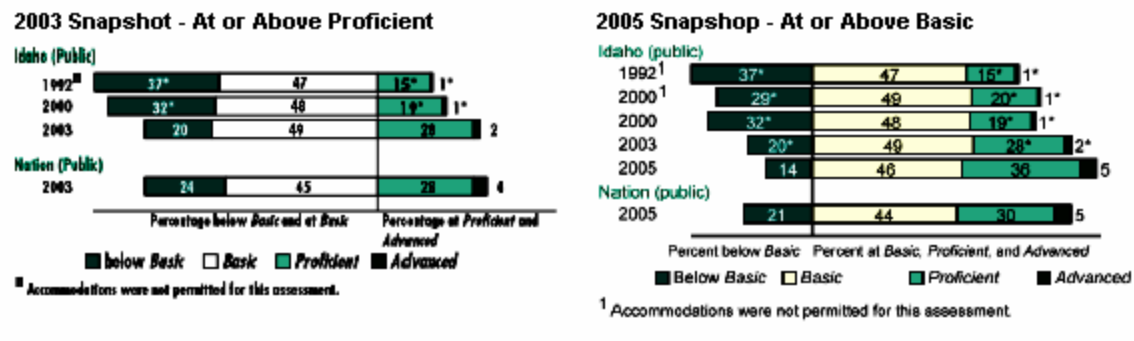


Figure 1. An illustration of changing the reporting focus from Percent At or Above *Proficient* to Percent At or Above *Basic* on NCES charts from the NAEP grade 4 Mathematics Snapshot Report for Idaho between 2003 and 2005.

An unforgivable digression. All of the claims about large differences in the reading proficiency percentages between state tests and NAEP in 2005 are based on comparing NAEP *Proficient* with state proficient. As Figure 2 illustrates, very different claims would have been made had NAEP *Basic* been compared with state proficient instead of NAEP *Proficient*. According to state and NAEP data provided by Hall and Kennedy (2006), point-by-point comparisons based on NAEP *Proficient* indicated that NAEP was more rigorous than the reading tests of 43 states while only one state test was more rigorous than NAEP (six states were missing data). On the other hand, comparisons based on NAEP *Basic* would have indicated that NAEP was more rigorous than the reading tests of only 11 states while 33 state tests were more rigorous than NAEP. The same interpretive pattern appeared in the corresponding mathematics comparisons. While this “point-by-point analysis” comparing state and NAEP proficiency scores in this digression is interesting, it should never have been done. Please forget that you saw it.

Point-by-point Analysis. An Ad Hoc committee convened by NAGB studied how the Secretary of Education might use NAEP scores to confirm state testing results as NCLB permitted. The committee recommended that, “‘Informed judgment’ 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).

Rigor of Middle School NAEP vs State Tests, 2005

Data Source: Hall and Kennedy, 2006.

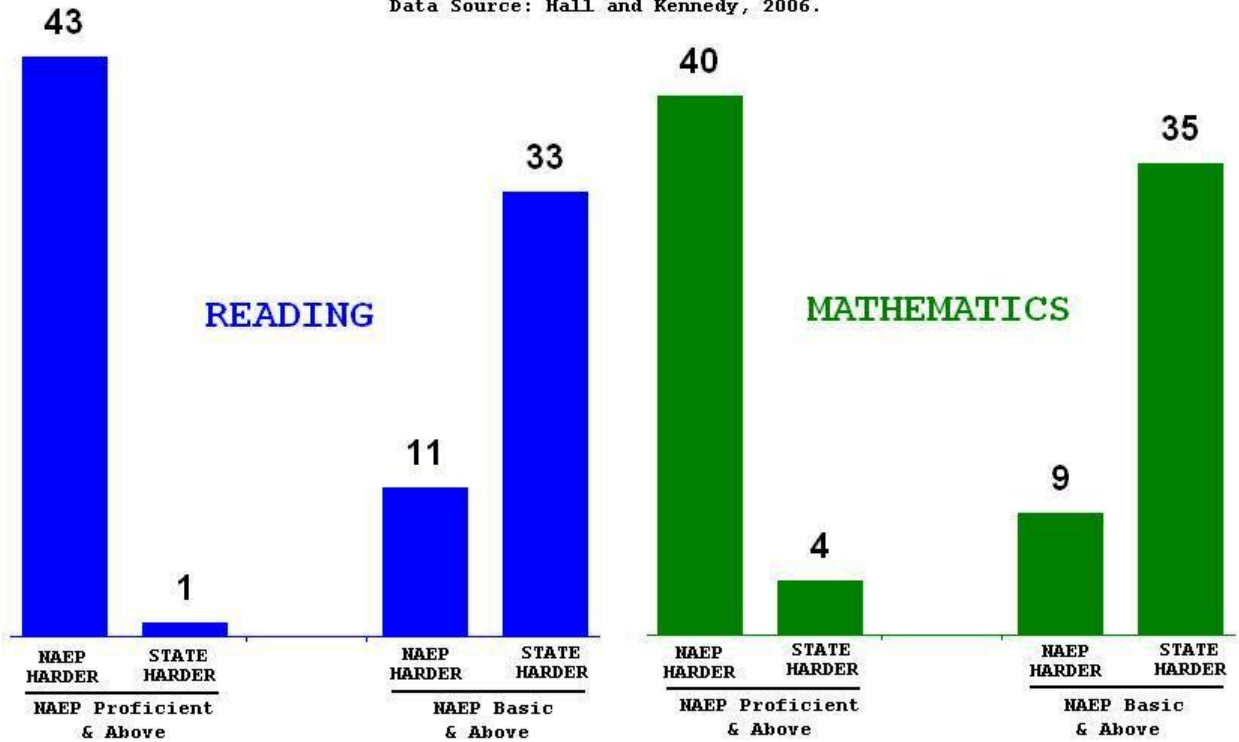


Figure 2. An illustration of how comparing NAEP *Basic* and above statistic rather than NAEP *Proficient* and above to the state proficiency statistic yields vastly different interpretations about the rigor of Idaho's 2005 state test results in reading and mathematics relative to NAEP.

The purpose of using NAEP is to provide a second “snapshot” of a state’s overall achievement. The Ad Hoc Committee highlighted several factors with the potential to limit the meaningfulness of a comparison between NAEP and state test results. In fact, these potential differences could even justify interpreting identical percentage scores from the state and NAEP tests as being quite different.

Potential differences between NAEP and state testing programs include: content coverage in the subjects, definitions of subgroups, changes in the demography within a state over time, sampling procedures, standard-setting approaches, reporting metrics, student motivation in taking the state test versus taking NAEP, mix of item formats, test difficulty, etc. Such differences may be minimal or great in number and in size and cannot reasonably be expected to operate in all states in equal fashion (Ad Hoc Committee, 2002).

The potential differences between NAEP and state testing programs rules against making any point-by-point analysis. However, differences between the testing

programs should be explored and reported even when other methods of analysis are used.

The Explanation . . .

The organizations and individuals who conducted the research and prepared the reports cited in this paper either (1) were unaware of the guidance for conducting a valid confirming analysis that is easily retrievable from the literature, or (2) chose to ignore that guidance. As a result, the methodologies employed in their studies were very much suspect and their findings were highly questionable if not out-and-out wrong.

How a Confirming Analysis Might Be Done.

The Ad Hoc Committee for Confirming Test Results--after preparing a long list of cautions and practices to avoid--did recommend one use of NAEP scores for confirming state test results. Indeed, NAEP achievement levels can be used as evidence to confirm the general trend of state test AYP results in grades 4 and 8 reading and mathematics (Ad Hoc Committee, 2002). This particular use coincides with the original purpose of NAEP, to measure student achievement and to report performance trends over time.

Congress has mandated external evaluations of NAEP, the most recent of which was conducted by the National Academy of Sciences. The Academy's findings regarding the standard-setting procedures and the use of NAEP achievement levels were extremely unfavorable. Nonetheless, the Academy did suggest that NAEP achievement levels might be used to report trends. "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" (Pellegrino, Jones, & Mitchell, 1998).

Figure 3 illustrates how NAEP might be used to confirm state testing results (Carr, 2002). It's a useful graphic for bringing together the points discussed in this paper. By comparing NAEP's percent at or above *Basic* to the state's percent at or above grade level (i.e., at or above proficient, in NCLB terms), the confirming analysis in Figure 3 recognizes that NAEP's definition of *Proficient* is not synonymous with grade-level proficiency in a subject. The different fill colors suggest differences between the two tests, which should be discussed in a narrative accompanying the graph. Moreover, the graph avoids point-by-point comparisons between NAEP and state achievement levels. Rather, it relies on the comparison of proficiency trend lines, a defensible method for using NAEP to confirm state AYP results.

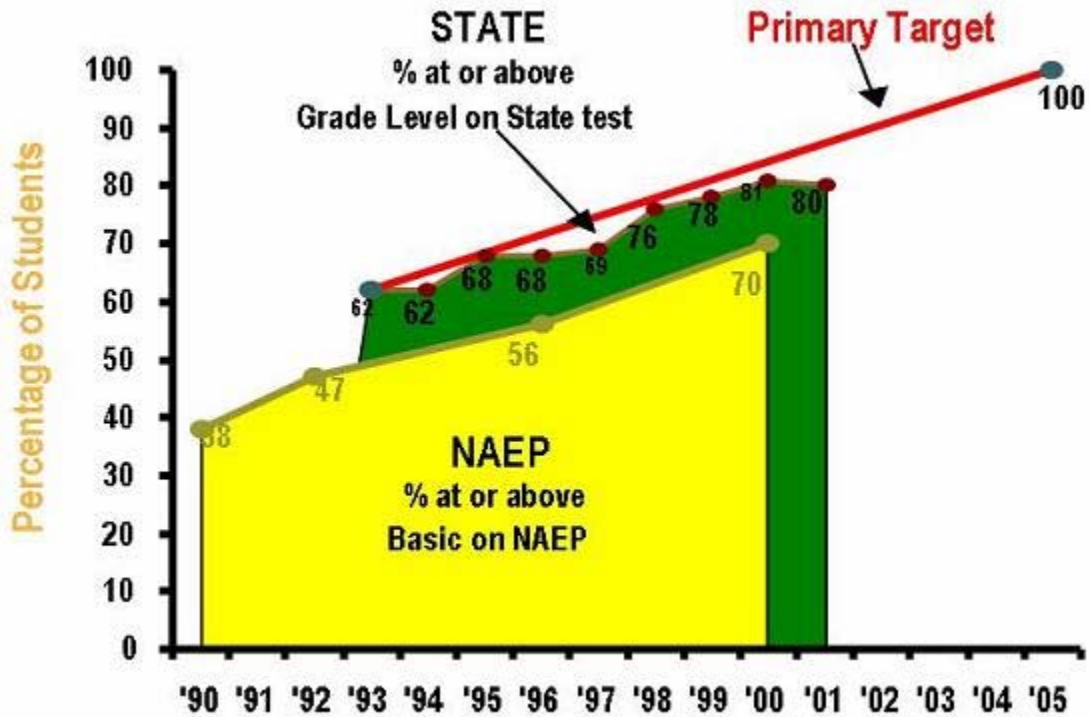


Figure 3. Graphic illustration of how NAEP percent at or above *Basic* as used pre-NCLB might be used to confirm state test results in the No Child Left Behind Act of 2001 (courtesy of Wendy Yen).

#

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.: U.S. Department of Education, National Assessment Governing Board. Retrieved March 27, 2007 from http://www.nagb.org/frameworks/reading_fw_03_07_prepub_edition.doc

Carr, P.G. (2002, August). *Legislative and Policy Update*. PowerPoint presentation at the National Assessment of Educational Progress (NAEP) State Coordinator Two-day Orientation of the NAEP State Service Center, Washington, D.C.

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.

Loomis, S.C., and Bourque, M.L. (Eds.) (2001). *National Assessment of Educational Progress achievement levels 1992-1998 for reading*. Washington, D.C.: U.S. Department of Education, National Assessment Governing Board. Retrieved December 4, 2006, from <http://www.nagb.org/pubs/readingbook.pdf>

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

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>

Reading framework for the 2005 National Assessment of Educational Progress. (2004) Washington, D.C.: U.S. Department of Education, National Assessment Governing Board. Retrieved December 4, 2006, from http://www.nagb.org/pubs/r_framework_05/761507-ReadingFramework.pdf

Stoneberg, B.D. (2007). Using NAEP to confirm state test results in the No Child Left Behind Act. *Practical Assessment, Research & Evaluation*, 12(5). Retrieved on June 1, 2007, from <http://pareonline.net/getvn.asp?v=12&n=5>

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. (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>

References (Promoting Claims of “Large Differences”)

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>

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

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>

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/op-ed/ravitch/20051107.htm>

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/b1164323.html>

Citation:

Stoneberg, Bert D. (2007, June). *An explanation for the large differences between state and NAEP proficiency scores reported for reading in 2005*. Paper presented at the Chief State School Officers (CCSSO) 37th Annual National Conference on Large-Scale Assessment, Nashville, TN.

**Council of Chief State School Officers
37th Annual National Conference on Large-Scale Assessment
Nashville, Tennessee
June 19, 2007**

An Explanation for the Large Differences between State and NAEP “Proficiency” Scores Reported for Reading in 2005



**Bert Stoneberg, Ph.D.
NAEP State Coordinator
Idaho State Board of Education
bert.stoneberg@osbe.idaho.gov**

HOW THIS STUDY STARTED ... EMAIL!

**>>> “Senator X <senate.idaho.gov>” 10/24/05 4:59 PM >>>
Dr. Howard – Is there an explanation for the
difference in Idaho and NAEP scores in reading as
the attached article says?**

>>> Marilyn Howard 10/24/05 5:58 PM >>>

Dear Senator X,

**I am forwarding your question to Dr. Stoneberg who
is the NAEP Testing Coordinator at the SDE. He
will respond to your question.**

Fordham Foundation Methodology (Leischer, 2005)

- ▶ **Assumed “state proficient” and “NAEP *Proficient*” have a common definition**
- ▶ **Computations**
 - Change in percent state proficient or above (AYP) from 2003 to 2005
 - Change in percent NAEP *Proficient* or above from 2003 to 2005
 - Point-by-point difference between the state and NAEP changes
- ▶ **Discovered large differences between changes in the state proficient and NAEP *Proficient* scores**
- ▶ **Attributed differences to states dumbing down their testing programs, actively participating in a “race to the bottom.”**
- ▶ **Rank ordered states according to their point-by-point difference between state and NAEP changes.**
- ▶ **Named the worst offenders (my state, Idaho, was included in the list).**

Other reports promoting the “large differences” claim:

Brookings Institution (Ravitch, 2005)

Center for American Progress (Rocha & Brown, 2005)

Education Trust (Hall & Kennedy, 2006)

Hoover Institution (Finn & Ravitch, 2006; Peterson & Hess, 2006)

National Center for Research on Evaluation, Standards, and Student Testing (Linn, Baker & Herman, 2005)

Northwest Regional Educational Laboratory (Greenough, 2005)

Policy Analysis for California Education (Fuller, Gesicki, Kang & Wright, 2006)

Rand Corporation (McCombs & Carrol, 2005).

U.S. Chamber of Commerce (Institute for a Competitive Workforce, 2007)

- **This list is illustrative, not exhaustive.** ●

Issues for a Confirming Analysis

- **Definition of “proficiency”**
 - NCLB definition required for state tests
 - NAGB definition for NAEP
- **Methodology for confirming analyses**
 - NAEP *Basic* and state proficient
 - Point-by-point comparisons
 - Differences between testing programs
 - How it might be done....

The U.S. Department of Education is responsible to say how NAEP scores are to be interpreted and used.



Professional standards for educational and psychological testing were set in 1999 by

▶ American Educational Research Association

▶ American Psychological Association

▶ National Council on Measurement in Education.

Valid Use of Test Scores

- ▶ **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.**

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. *Standards for educational and psychological Testing*. Washington, D.C.: American Educational Research Association, 1999.

Standards under this paragraph shall...

“(II) describe two levels of high achievement (proficient and advanced) that determine how well children are mastering the material in the State academic content standards;”

No Child Left Behind Act, 2001

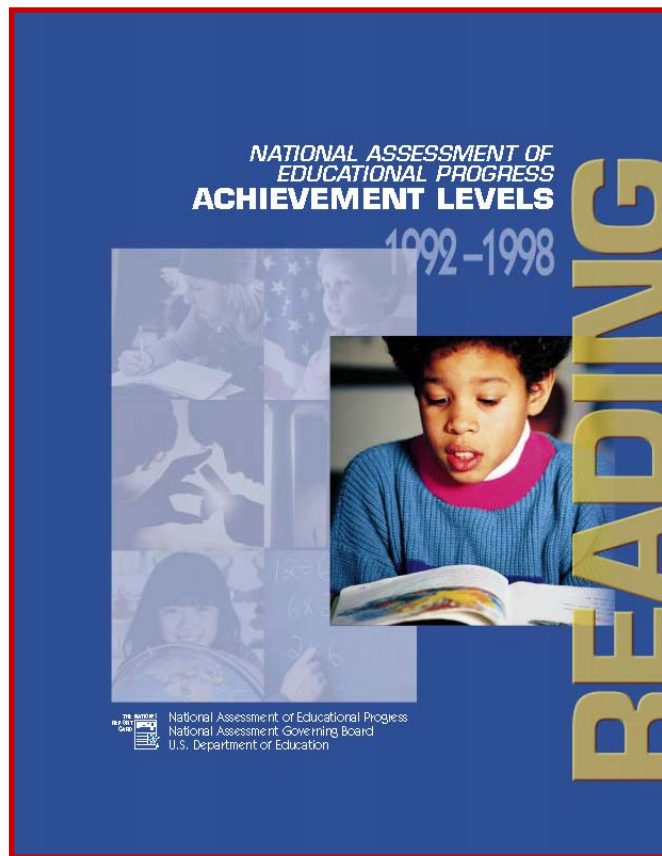
The “proficient” achievement level represents attainment of grade-level expectations for that academic content area.

Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001. Washington, D.C.: U.S. Department of Education, 2004.

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.

Building on Results: A Blueprint for Strengthening the No Child Left Behind Act. Washington, D.C.: U.S. Department of Education, 2007.

NAEP's definition of "*Proficient*" is not bound by grade-level expectations or proficiency in a subject.



Achievement Levels Report

Reading

**Writing, Mathematics,
Science, U.S. History,
Geography, and Civics**

Loomis, S.C., and Bourque, M.L. (Eds.). *National Assessment of Educational Progress Achievement Levels, 1992-1998 for Reading*. Washington, DC: National Assessment Governing Board, 2001.

How Should Achievement Levels Be Interpreted?

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.

Finally, even the best students you know may not meet the requirements for Advanced performance on NAEP.

**Reading Framework for the
2005 National Assessment
of Educational Progress**



National Assessment Governing Board
U.S. Department of Education

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.

Reading Framework for 2005, 2004.

READING FRAMEWORK

for the

2009 National Assessment of Educational Progress

PRE-PUBLICATION EDITION

2007

Prepared for the

National Assessment Governing Board
In support of Contract No. ED-02-R-0007

American Institutes for Research
1000 Thomas Jefferson Street, N.W.
Washington, DC 20007

Proficient readers will have sizeable meaning vocabularies, including knowledge of many words and terms above grade level.

Reading Framework for the 2009 National Assessment of Educational Progress (Pre-publication Edition), 2007.

Achievement Level	NAEP Achievement Level Descriptors	Letter Grade Range
Advanced		A+
Proficient	Some of the best students you know Many words and terms above grade level Mastery	A ↕ B+
Basic	Proficiency in subject (common meaning) Overall understanding of grade-appropriate text More than minimal competency	B ↕ C-
Below Basic	Minimally competent	D+ ↕ F

Descriptors and estimated letter grade ranges for NAEP achievement levels.

NAEP percent at or above *Basic* is the most directly comparable statistic for confirming state AYP results.

Mosquin, P., and Chromy J. *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, 2004

Federal Sample Sizes for Confirmation of State Tests in the No Child Left Behind Act

Paul Mosquin
RTI International

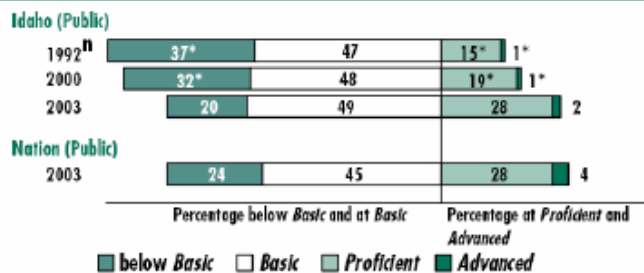
James Chromy
RTI International

Commissioned by the NAEP Validity Studies (NVS) Panel
May 2004

George W. Bohrnstedt, Panel Chair
Frances B. Stancavage, Project Director

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

Student Percentage at NAEP Achievement Levels

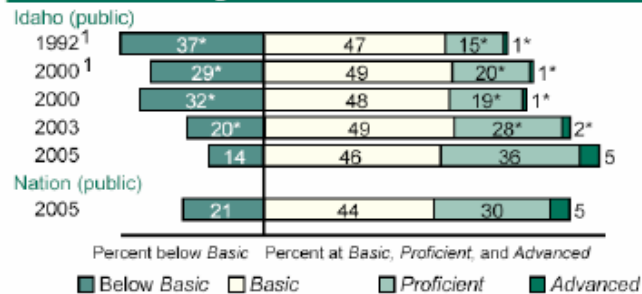


ⁿ Accommodations were not permitted for this assessment.

NOTE: The NAEP mathematics scale ranges from 0 to 500, with the achievement levels corresponding to the following points: *Below Basic*, 213 or lower; *Basic*, 214-248; *Proficient*, 249-281; *Advanced*, 282 or above.

Idaho Snapshot Report
Mathematics 2003, Gr 4

Student Percentage at NAEP Achievement Levels



¹ Accommodations were not permitted for this assessment.

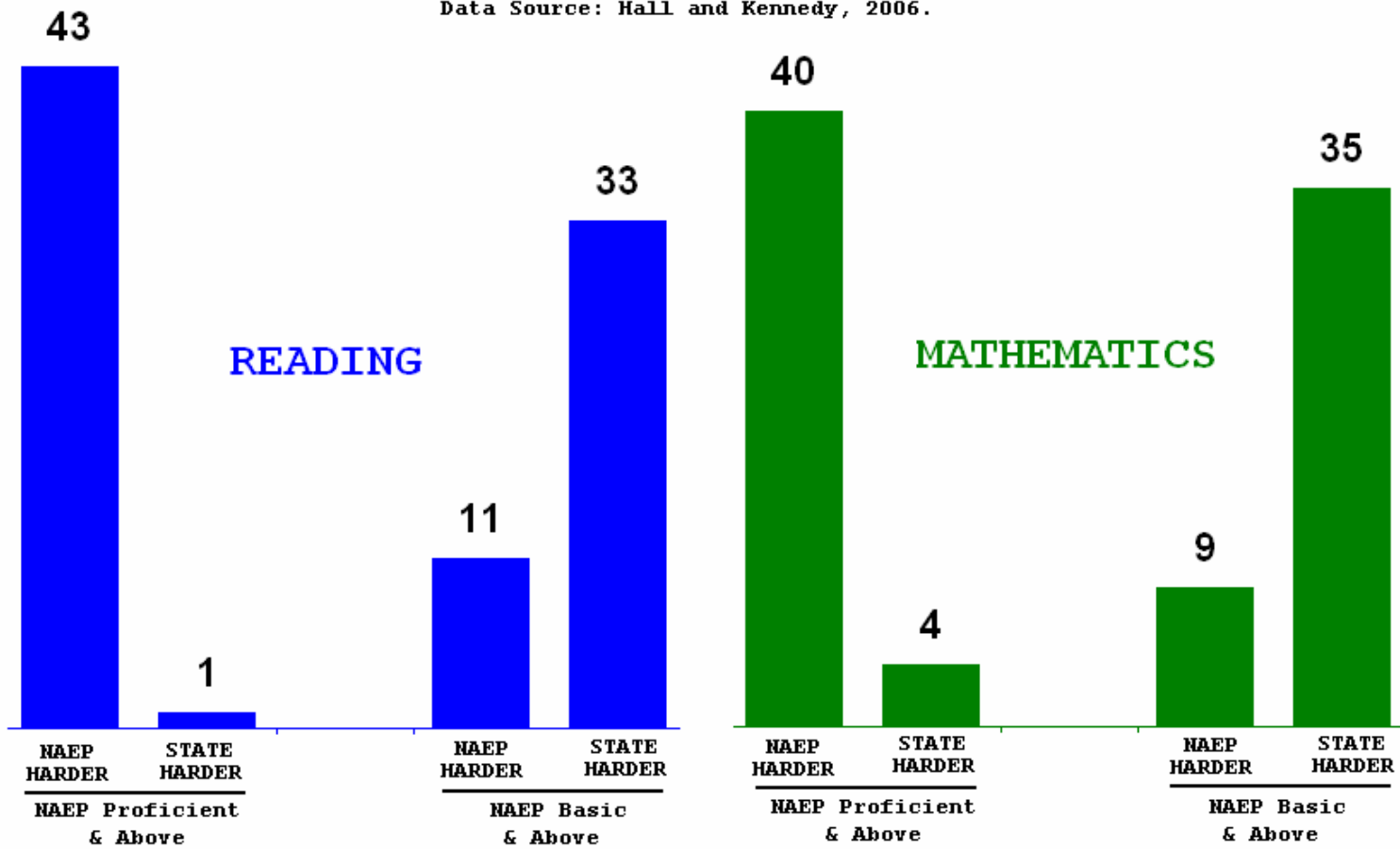
NOTE: The NAEP mathematics achievement levels correspond to the following scale points: *Below Basic*, 213 or lower; *Basic*, 214-248; *Proficient*, 249-281; *Advanced*, 282 or above.

Idaho Snapshot Report
Mathematics 2005, Gr 4

Note: In some NCES prepared reports with results from NAEP 2005, the percent at or above *Basic* was given prominence for the first time. This change in reporting practice is in harmony with the NAEP Validity Studies Panel's recommendations.

Rigor of Middle School NAEP vs State Tests, 2005

Data Source: Hall and Kennedy, 2006.



National Assessment Governing Board
National Assessment of Educational Progress

**Using the National Assessment of Educational Progress
To Confirm State Test Results**

A Report of
The Ad Hoc Committee on Confirming Test Results

March 1, 2002

Ad Hoc Committee on Confirming Test Results
Michael Nettles, Chair
Daniel Domenech
Edward Haertel
Nancy Kopp
Debra Paulson
Diane Ravitch
Michael Ward
Marilyn Whirry
Dennie Palmer Wolf

Planning Work Group
Mark Reckase, Chair
Peter Behuniak
David Francis
Paul Holland
Scott Jenkins
Mary Jean LeTendre
Gerry Shelton
Wendy Yen

Governing Board Staff
Ray Fields

**Confirmation of
state AYP results
should NOT be
conducted on a
point-by-point basis.**

**When confirming
state AYP results,
differences between
NAEP and the state
testing program
must be explored
and reported.**

“Potential differences between NAEP and state testing programs include: content coverage in the subjects, definitions of subgroups, changes in the demography within a state over time, sampling procedures, standard-setting approaches, reporting metrics, student motivation in taking the state test versus taking NAEP, mix of item formats, test difficulty, etc. Such differences may be minimal or great in number and in size and cannot reasonably be expected to operate in all states in equal fashion.”

Ad Hoc Committee on Confirming Test Results. *Using the National Assessment of Educational Progress to confirm state test results*. Washington, D.C.: National Assessment Governing Board, 2002.

So, the explanation for the “large differences” between state and NAEP reading scores in 2005?

The researchers and organizations responsible for reporting the “large differences” were either

(1) Unaware of information about NCLB and NAEP related to conducting a valid “confirming analysis” that is readily available in public documents, or

(2) Chose to ignore that information.

As a result, their methodology was flawed and their claims about “large differences” were highly questionable, if not out-and-out wrong.

How it might be done...

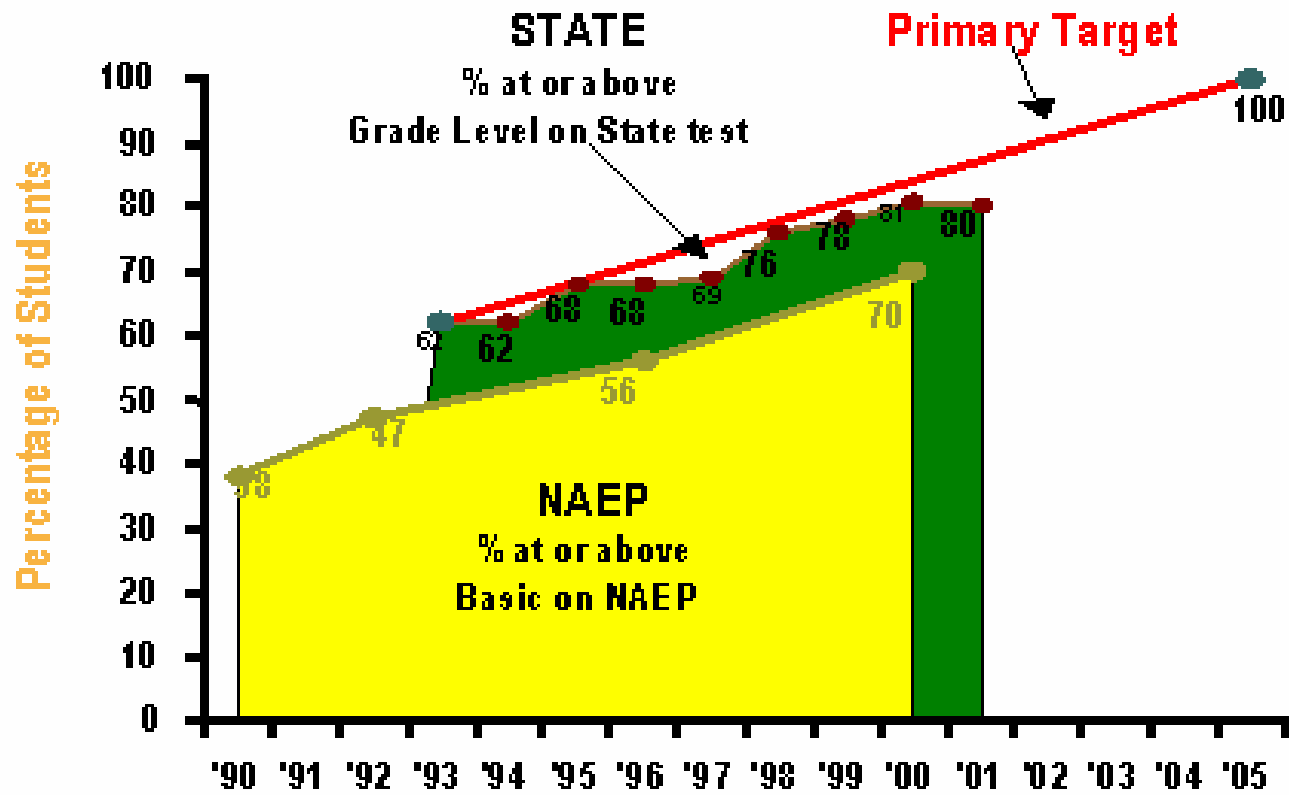
NAEP can be used as evidence to confirm the general trend of state test results in grades 4 and 8 reading and mathematics.

Ad Hoc Committee on Confirming Test Results. (2002).

“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....”

Pellegrino, J.W., Jones, L.R., and Mitchell, K.J. (Eds.). *Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press, 1998.

How it might be done...



Carr, P.G. (2002, August). *Legislative and Policy Update*. PowerPoint presentation at the National Assessment of Educational Progress (NAEP) State Coordinator Two-day Orientation of the NAEP State Service Center, Washington, D.C.

For additional information see...

Stoneberg, B.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>

