

NAEP State Service Center  
Spring Assessment Literacy Workshop  
DoubleTree Hotel – Bethesda, Maryland  
March 13, 2009

## **A Quick Review of NCLB (State) and NAEP Achievement Levels and How They Match**

Bert D. Stoneberg  
Idaho State Board of Education

### **Abstract**

Test developers are responsible to define how test scores should be interpreted and used. The No Child Left Behind Act of 2001 (NCLB) directed the Secretary of Education to use results from the National Assessment of Educational Progress (NAEP) to confirm the proficiency scores from state developed tests. There are two sets of federal definitions for the term “proficient,” one NAEP and one for NCLB. NAEP’s “At or Above *Basic*” is the most directly comparable statistic for confirming state proficiency results. NAEP and state proficiency scores, however, should be used (and interpreted) with caution. Achievement level results may provide useful trend information for one group on one test, but the statistical properties of proficiency scores render them ill-suited for trend comparisons. It may well be that there is no defensible, statistical method for using NAEP achievement level results to confirm a state’s proficiency scores. Until the federal law is amended proficiency score analyses it requires should be accompanied, whenever possible, by related analyses based on scale scores or effect size or both.

### **Introduction**

This workshop session revisits a paper the author presented at the 2007 national conference on large-scale assessment entitled “An Explanation for the Large Differences between State and NAEP Proficiency Scores Reported for Reading in 2005” (Stoneberg, 2007a). The content for the beginning and middle of today’s presentation has much in common with that paper, but the ending is remarkably different.

## **Standards for Educational Testing**

The standards for educational and psychological testing – jointly established by the American Educational Research Association, the American Psychological Association, and the National Council for Measurement in Education – address the valid use of test scores (Joint Committee on Standards, 1999). Standard 1.2, for example, says “the test developer should set forth clearly how test scores are intended to be interpreted and used.” Standard 1.4 says “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.”

The National Assessment Governing Board (NAGB) sets policy for NAEP while the National Center for Education Statistics (NCES) implements it. NAGB and NCES together constitute the “test developer” for NAEP. The state is the “test developer” for the state test, but the state must abide by federal statute and regulation as guided by NCLB program officials in U.S. Department of Education.

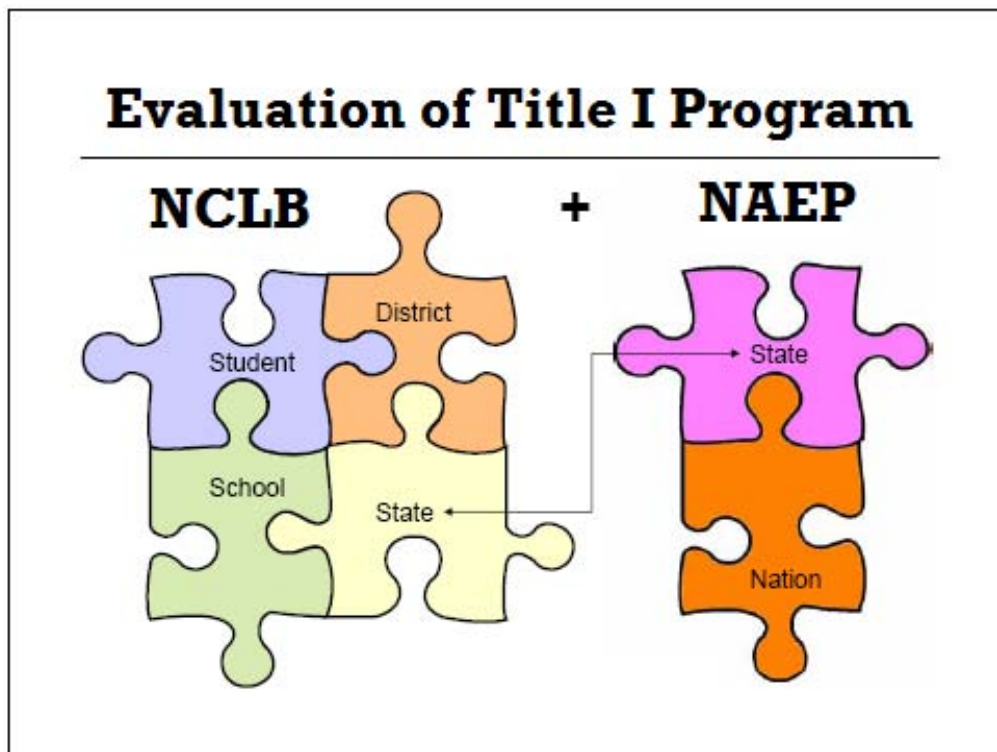
### **A New Use for NAEP**

Since its creation in 1969, NAEP has had two major goals: to assess student performance reflecting current educational and assessment practices, and to measure change in student performance reliably over time. To this end, NAEP has given careful attention to the standards for educational and psychological testing as established by the community of professionals engaged in educational research, measurement and evaluation, psychometrics, and statistics.

The No Child Left Behind Act of 2001 (NCLB) created a new use for NAEP by stipulating that “the Secretary shall use information from a variety of sources, including the National Assessment of Educational Progress [...], state evaluations, and other research

studies” to assess or evaluate the Title I program. The apparent motivation for requiring NAEP was to keep the states honest through external confirmation of the results that states reported for their NCLB tests. NAEP would serve this purpose well because all states would participate in NAEP and no state would have any control over the national assessment. Figure 1 illustrates the levels of scores available from the state NCLB test (i.e., student, school, district and state) and NAEP (state and national). The challenge has been to come up with a defensible procedure to compare or match the state-level results from the NCLB test and NAEP.

**Figure 1. Levels of results reported for the NCLB (state) test and NAEP. The challenge is to match state level results from both tests.**



NCLB required a state to develop its assessment so it could report on “two levels of high achievement (proficient and advanced) that determine how well children are mastering the material in the State academic content standards.” NCLB placed focus on

reporting out the percentage of students scoring at or above proficient on the state test. This statistic is known as AYP or Adequate Yearly Progress. It is unfortunate, but the narrow focus on the use of achievement level results in NCLB rendered the use of NAEP achievement level scores to confirm state AYP reports unavoidable. NCLB and NAEP use the same “names” for the various achievement levels (i.e., basic, proficient, and advanced), but the NCLB-mandated state tests and NAEP operate under different definitions for each achievement level name. It is a mistake to assume that proficient is proficient is proficient. NCLB’s state proficient is not the same as *NAEP Proficient*.

### **NCLB Achievement Levels: Interpretation and Use**

The U.S. Department of Education implemented a peer review process to provide federal oversight as states developed their NCLB tests. A peer review team made up of out-of-state persons with expert knowledge and skills in curriculum and assessment visited the program. The team filled out an extensive review checklist while on-site, and issued a report with findings and recommendations. The Title I programs in some states were fined because they did not corrective action sufficient to “pass” peer review on subsequent visits.

The peer review team was required to examine the state’s definitions for the achievement levels. In particular, the team had to pass judgment on the state’s definition of proficient. It had to mark Yes or No on the checklist whether “The ‘proficient’ achievement level represents the attainment of grade-level expectations for that academic content area.” (U.S. Department of Education, 2004).

It’s noteworthy here that before NCLB some state testing programs used out-of-level testing for students whose instructional levels were either below or above their grade level. This practice, however, did not survive the peer review process. It was made clear that

NCLB required state tests to measure achievement of grade-level content and to be administered to students at that grade. “On-grade-level, at-grade-level” became the mantra for state tests under NCLB.

While preparing for the reauthorization battle over the No Child Left Behind Act, the U.S. Department of Education published its blueprint for strengthening the law. It said, “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). Substituting achievement level names specified in NCLB, this might be taken to mean that the intent of the law was to ensure that all students could read and do math at the proficient level or the advanced level. Indeed, the Department’s blueprint made it clear that state “proficient” means “at grade level” and that advanced means “better than grade level.”

### **NAEP Achievement Levels: Interpretation and Use**

The National Assessment Governing Board (NAGB) has not been silent about the interpretation and use of NAEP achievement level scores. It published achievement level reports to explain its interpretation of achievement level scores. The Board convened an Ad Hoc committee to study how NAEP might be used to confirm state test results, and received reports from the NAEP Validity Studies Panel. It has also published a framework for each assessment that expands upon the policy definitions of *Basic*, *Proficient*, and *Advanced*.

Achievement Level Reports. In 2001, as Bush and Kennedy lead the passage of the No Child Left Behind Act, NAGB published a series of booklets to inform the public about the interpretation and use of NAEP scores. Text from the reading booklet (identical

language is also found in the booklets for writing, mathematics, science, U.S. history, geography, and civics) reads:

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).

Ad Hoc Committee Report. In 2002, the Board’s Ad Hoc Committee on Confirming Test results issued its report. The committee’s work did not examine (i.e., compare or contrast) the differing interpretations and uses that NCLB and NAEP had stipulated regarding the achievement levels, whether basic, proficient, or advanced. The report, however, did contain several important findings. Three key findings to consider. First, NAEP can be used as evidence to confirm the general trend of state test results in grades 4 and 8 reading and mathematics. Second, confirmation of state AYP results should NOT be conducted on a point-by-point basis. Third, when confirming state AYP results, differences between NAEP and the state testing program must be explored and reported. (Ad Hoc Committee, 2002).

NAEP Validity Studies Panel. In 2004, the NAEP Validity Studies Panel issued a report for a statistical analysis that concluded “NAEP’s ‘percent At or Above *Basic*’ is the most directly comparable statistic for confirming state AYP results.” (Mosquin & Chromy, 2004). When results from NAEP 2005 were released, the percent at or above *Basic* was given prominence in some reports for the first time ever. The NAEP reports were prepared

by the National Center for Education Statistics and released by the National Assessment Governing Board. This change in reporting practice indicated that both parties accepted the NAEP Validity Studies Panel's findings as consistent with existing NAEP policy and practice.

NAEP Frameworks. NAEP frameworks are not curriculum documents that express what students should be learning in America's schools. They are a description of what will be tested and how the scores should be interpreted. The framework for each subject expands NAEP's policy definitions of *Basic*, *Proficient*, and *Advanced*.

The policy definition notes that *Basic* "denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade." Language from the framework for the 2007 reading assessment clarifies "prerequisite knowledge and skills" for *Basic* at the fourth grade. 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. (National Assessment Governing Board, 2006).

Language from the framework for the NAEP 2009 reading assessment clarifies "prerequisite knowledge and skills" for *Proficient*. "Proficient readers," it says, "will have sizeable meaning vocabularies, including knowledge of many words and terms above grade level." (National Assessment Governing Board, 2008). This contrasts with NCLB's "on-grade-level, at-grade-level" yoke.

Clearly, this language from the NAEP reading frameworks indicates that NAEP *Basic* represents an estimate of "grade-level expectations," and that NAEP *Proficient* demands some above-grade-level knowledge and skills. Once again, NCLB requires a state to define

proficient as meeting grade-level expectations on state content. There can be no doubt that using a state’s NAEP *Proficient* score to confirm a state’s NCLB proficient score would surely result in mistaken and misleading conclusions. This, however, has been the prime methodology over the last half-decade for NAEP-state proficiency analyses conducted and published by national foundations, institutes, and think tanks.

**NAEP Achievement Levels and “Letter-Grades”**

One way to understand the NAEP achievement levels is to link NAEP’s descriptive language to letter grades (i.e., A, B, etc.) that one would likely see on the report cards of students performing at each NAEP achievement level (Stoneberg, 2007b). Figure 2 compares the language used to describe NAEP achievement level scores and “letter grades” used to describe corresponding classroom performance levels. The language describing NAEP *Basic* corresponds letter grades ranging from C- to B, which represents meeting grade-level expectations for that particular grade.

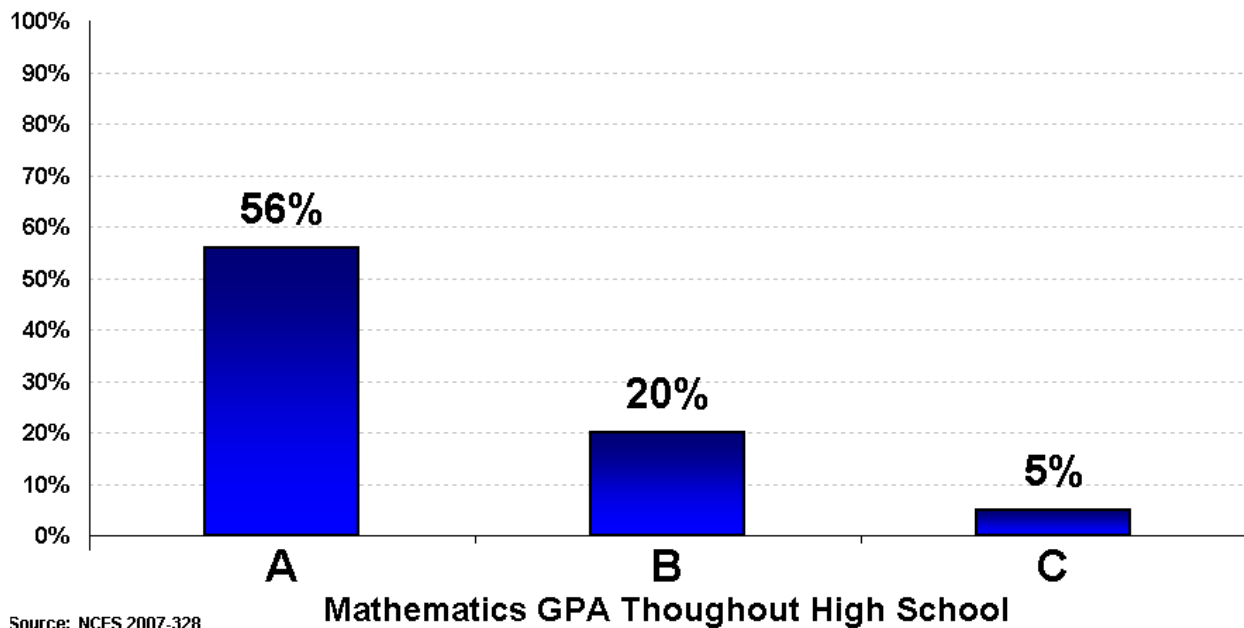
**Figure 2. Comparing language used to describe the NAEP achievement levels and “letter grades” used to describe corresponding classroom performance.**

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



In 2007, NCES published a statistical analysis report finding that “A majority (56 %) of Proficient and above performers on the 1992 NAEP-scaled mathematics assessment maintained an “A” average in mathematics throughout high school. Some 20 percent of “B” students and 5 percent of “C” students reached the proficient or advanced levels.” (Scott & Ingles, 2007). Figure 3 presents these results graphically. The interpretation of these results are muddled somewhat because not all students take the same mathematics course in high school. One student may have an A average in two courses (e.g., general math and consumer math, really 8<sup>th</sup> grade arithmetic a second time and a third time), while another student may have an A average in four rigorous courses including AP Statistics and Math Analysis. The latter will likely reach *Proficient* on NAEP, while the former probably will not. However, a student with a C average through four rigorous mathematics courses may still reach the NAEP *Proficient* level. In general, these results leave the impression that NAEP *Proficient* requires a performance that is higher than just meeting grade-level expectations.

**Figure 3. The percentage of high school seniors by mathematics GPA who scored “At or Above Proficient” on a 1992 NAEP-scaled mathematics assessment.**



Source: NCES 2007-328

## **Use Achievement Level Scores with Caution**

Congress has mandated external evaluations of NAEP, the most recent of which by the National Academy of Sciences (Pellegrino, Jones & Mitchell, 1998). The Academy found that NAEP's procedure for setting cut-scores was fundamentally flawed because it rested on "informed judgment" rather than a "highly objective process," and noted that the process had produced some unreasonable results. Even though its report was highly critical of NAEP's achievement levels, the Academy did recommend their cautious use for drawing attention to changes in student performance over time.

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 [...] the reports should strongly and clearly emphasize that 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 [...] rather than focusing on the percentages in each category in a single year. (Pellegrino, Jones & Mitchell, 1998).

In NCLB, Congress required the Secretary to use NAEP data to evaluate the Title I program, but NCLB also required that NAEP achievement levels be used on a trial basis until the Commissioner of Education Statistics determines that the achievement levels are "reasonable, valid, and informative to the public." Until that determination is made, the law requires the Commissioner and the National Assessment Governing Board to state clearly the trial status of the achievement levels in all NAEP reports.

The website for the "Nation's Report Card" notes that, "The Board and NCES believe that the achievement levels are useful for reporting trends in the educational achievement of students in the United States. However, [...] NCES concludes that these achievement levels should continue to be used on a trial basis and should continue to be interpreted and

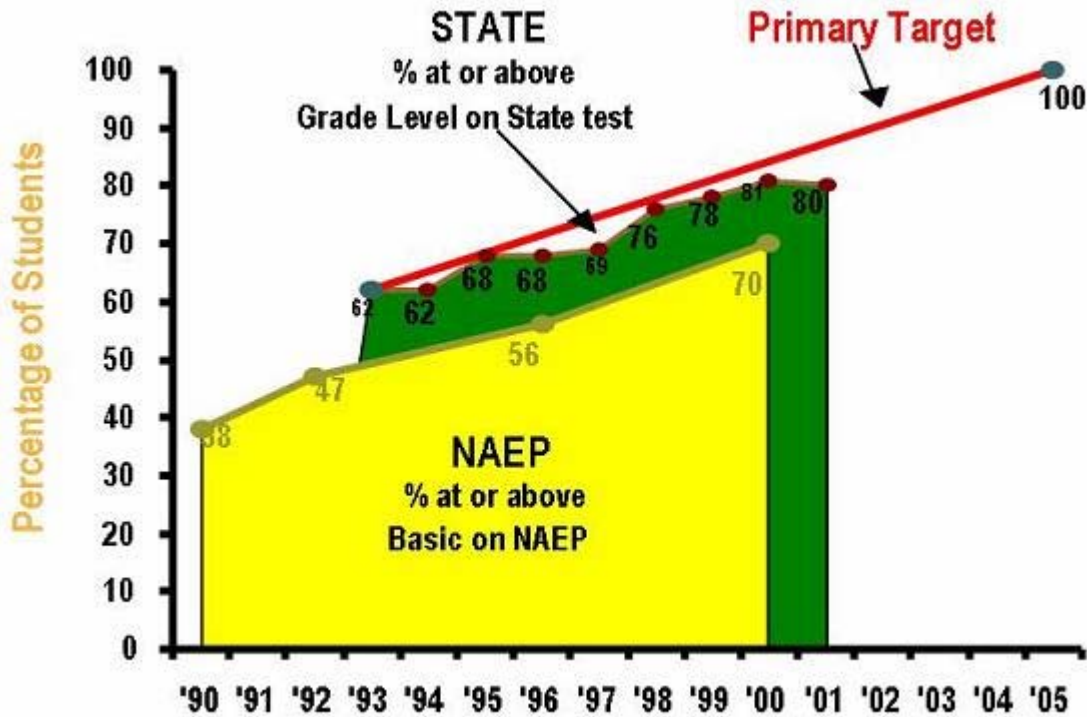
used with caution.” The Board and NCES also note on the website about the Nations Report Card that a proven alternative to the current process of setting cut-scores has not yet been identified. They invite organizations and individuals with ideas for alternative models for setting cut-scores to present them for consideration.

### **How Might a Confirming Analysis Be Done? Then.**

Given the interpretation and status of NAEP achievement levels and the stated purpose of the national assessment, it seemed in 2007 to the author that graphing trend lines plotting state percent at or above proficient and NAEP percent At or Above *Basic* side-by-side together offered a defensible method for confirming state AYP results. If the trend lines moved in the same direction, it indicated that NAEP confirmed the state results. At least, this was the notion advanced in the author’s paper presented at the CCSSO 37th Annual National Conference on Large-Scale Assessment in Nashville. Quote, complete with graphic (Stoneberg, 2007a):

Figure “A” 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 “A” 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 "A". Graphic illustration of how NAEP percent at or above *Basic* might be used to confirm state test results in the No Child Left Behind Act of 2001



**How Might a Confirming Analysis Be Done? Now!**

The defensible method for conducting a NAEP-state confirming analysis by comparing their achievement level trends that the author presented at the Large-Scale Assessment Conference in Nashville in June 2007 was essentially rendered *indefensible* in December 2007.

In its 1998 evaluation report, the National Academy of Sciences did recommend that NAEP achievement level results might be used (with caution) to plot a performance trend for a group. Under the blanket of the Academy’s recommendation, it had been assumed generally that proficiency data from NAEP and NCLB tests enjoyed the requisite statistical properties for sound trend comparisons. An unexamined assumption!

A pivotal study by Andrew Ho (University of Iowa) that compared NAEP and state proficiency trend data, however, disputed the assumption. “Trend comparisons require both technical care and substantive consideration. As useful as PAC [percent above cut-score] statistics have been in communicating test results to the public, their properties as trend statistics render them ill-suited for trend comparison” (Ho, 2007).

Dr. Ho presented two sessions at this workshop yesterday related to proficiency standards and defensible methods for making NAEP-state comparisons. Four points from his presentations were particularly noteworthy:

- For NAEP-State comparisons, we need to get past proficiency standards.
- The proficiency metric distorts just about every important large-scale test-driven inference.
- Trend and gap interpretations can be inflated or deflated by cut-score location.
- High-stakes trends, gaps, and gap trends should all be reported on a scale-score or effect-size metric. (Ho, 2009).

The need to change metrics that Ho has advanced seems both credible and desirable. NAEP-state comparisons based on achievement level scores are indeed *per se* faulty. Unfortunately, the language in NCLB requires the Secretary to use the proficiency metric for NAEP-state comparisons. So until the federal law is changed, any analysis based on the proficient metric that might be required by NCLB should, whenever possible, be associated with and accompanied by a related analysis based on scale-scores or on effect sizes or on both scale scores and effect sizes.

# # #

## 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 March 7, 2009, from [http://www.nagb.org/publications/color\\_document.pdf](http://www.nagb.org/publications/color_document.pdf)

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

Ho, A.D. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice*, 26(4), pp. 11-20.

Ho, A.D. (2009, March). *NAEP-to-State Comparisons from 2003 to 2007: An Overview of Defensible Methods and Interpretations*. PowerPoint presentation at the NAEP State Service Center's Spring Assessment Literacy Workshop, Bethesda, MD.

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.: National Assessment Governing Board. Retrieved March 7, 2009, from <http://www.nagb.org/publications/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 March 7, 2009, from: [http://www.air.org/publications/documents/MosquinChromy\\_AIR1.pdf](http://www.air.org/publications/documents/MosquinChromy_AIR1.pdf)

National Assessment Governing Board. (2006). *Reading framework for the 2007 National Assessment of Educational Progress*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved March 7, 2009, from [http://nagb.org/publications/frameworks/reading\\_07.pdf](http://nagb.org/publications/frameworks/reading_07.pdf)

National Assessment Governing Board. (2008). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics. Retrieved March 7, 2007, from <http://nagb.org/publications/frameworks/reading09.pdf>

Scott, L.A., and Ingels, S.J. (2007). *Interpreting 12th-graders' NAEP-scaled mathematics performance using high school predictors and postsecondary outcomes from the National Education Longitudinal Study of 1988 (NELS:88)* (NCES 2007-328). Washington, D.C.: National Center for Education Statistics, U.S. Department of Education. Retrieved March 7, 2009, from <http://nces.ed.gov/pubs2007/2007328.pdf>

Stoneberg, B.D. (2007a). *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) National Conference on Large-Scale Assessment, Nashville, TN. Retrieved March 7, 2009, from <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED497395>

Stoneberg, B.D. (2007b). Using NAEP to confirm state testing results in the No Child Left Behind Act. *Practical Assessment, Research & Evaluation*, 12(5). Available online: <http://www.pareonline.net/pdf/v12n5.pdf>

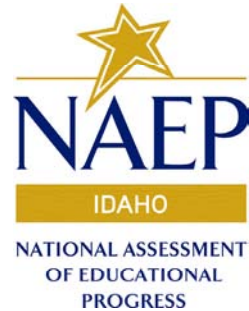
U.S. Department of Education. (2004). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*. Washington, D.C.: Author. Retrieved March 7, 2009, from <http://www.ed.gov/policy/elsec/guid/saaprguidance.doc>

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 7, 2009, from <http://www.ed.gov/policy/elsec/leg/nclb/buildingonresults.pdf>

### **Suggested Citation**

Stoneberg, B. D. (2009, March). *A Quick Review of NCLB (State) and NAEP Achievement Levels and How They Match*. Paper presented at the NAEP State Service Center (NSSC) Spring Assessment Literacy Workshop, Bethesda, MD.





---

# A Quick Review of NCLB (State) and NAEP Achievement Levels and How They Match

---

**Bert Stoneberg**  
**NAEP State Coordinator**  
**Idaho State Board of Education**  
<http://www.boardofed.idaho.gov/naep/>

**NAEP State Service Center**  
**Spring Assessment Literacy Workshop**  
**DoubleTree Hotel – Bethesda**  
**March 13, 2009**





**Standards for educational and psychological testing were updated in 1999 by a joint effort of the**

**▶ American Educational Research Association**

**▶ American Psychological Association**

**▶ National Council on Measurement in Education.**

## **Standards: 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.

## SUPPLEMENTAL NOTES

[<< Go Back](#)

### Note 4: National Assessment of Educational Progress (NAEP)

The National Assessment of Educational Progress (NAEP), governed by the National Assessment Governing Board (NAGB), is administered regularly in a number of academic subjects. Since its creation in 1969, NAEP has had two major goals: to assess student performance reflecting current educational and assessment practices and to measure change in student performance reliably over time. To address these goals, NAEP includes a main assessment and a long-term trend assessment. The two assessments are administered to separate samples of students at separate times, use separate instruments, and measure different educational content. Thus, results from the two assessments should not be compared.

**Since its creation in 1969, NAEP has had two major goals: to assess student performance reflecting current educational and assessment practices, and to measure change in student performance reliably over time.**

background questionnaires (for the student, teacher, and school) to provide information on instructional experiences and the school environment at each grade.

Public Law 107–110  
107th Congress

An Act

To close the achievement gap with accountability, flexibility, and choice, so that  
no child is left behind.

Jan. 8, 2002

[H.R. 1]

*Be it enacted by the Senate and House of Representatives of  
the United States of America in Congress assembled,*

**SECTION 1. SHORT TITLE.**

This title may be cited as the “No Child Left Behind Act  
of 2001”.

**“PART E—NATIONAL ASSESSMENT OF TITLE I**

20 USC 6491.

**“SEC. 1501. EVALUATIONS.**

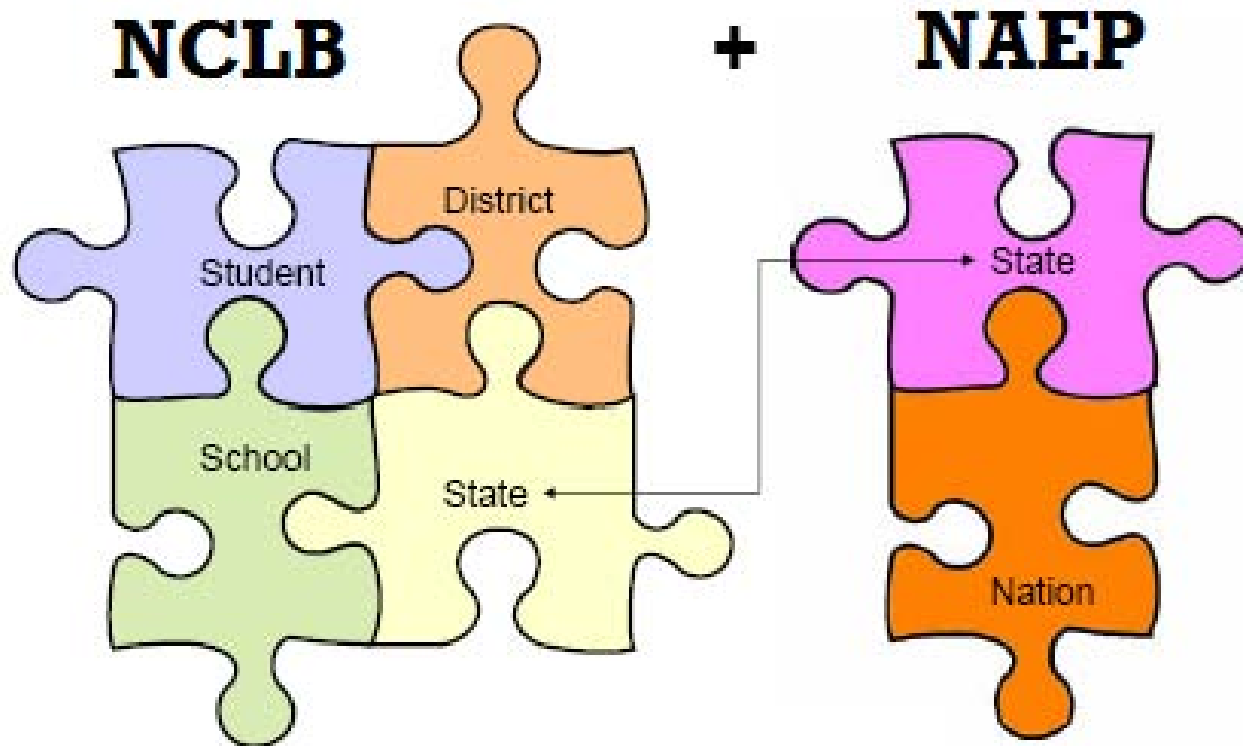
“(a) NATIONAL ASSESSMENT OF TITLE I.—

“(1) IN GENERAL.—The Secretary shall conduct a national  
assessment of the programs assisted under this title and the  
impact of this title on States, local educational agencies, schools,  
and students.

“(3) SOURCES OF INFORMATION.—In conducting the assess-  
ment under this subsection, the Secretary shall use information  
from a variety of sources, including the National Assessment  
of Educational Progress (carried out under section 411 of the  
National Education Statistics Act of 1994), State evaluations,  
and other research studies.

# Evaluation of Title I Program

---



## **No Child Left Behind Act of 2001**

**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;”**

# Interpretation and Use of NCLB Achievement Levels

## Office of Elementary and Secondary Education

**The “proficient” achievement level represents the 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.

## Secretary of Education

**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.

## **Interpretation and Use of NAEP Achievement Levels**

### **National Assessment Governing Board (NAGB)**

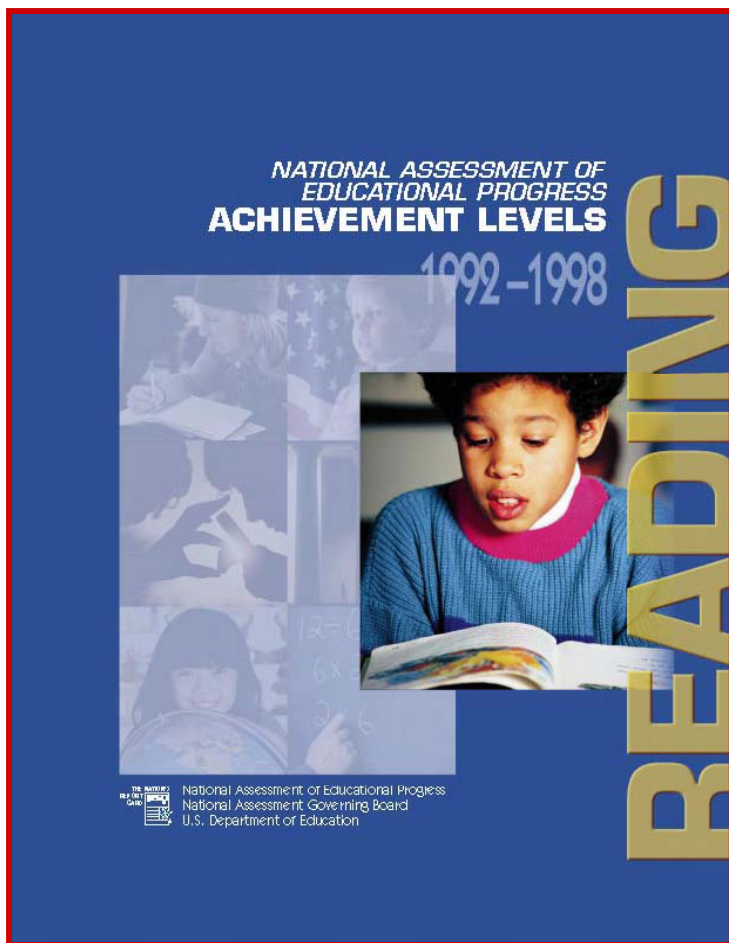
**NAEP Achievement Level Reports (2001)**

**NAGB Ad Hoc Committee Report (2002)**

**NAEP Validity Studies Panel Report (2004)**

**NAEP Frameworks**





# Achievement Levels Report

## Reading

---

---

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

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

## **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.**

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

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

**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.**

## **Differences 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.

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*

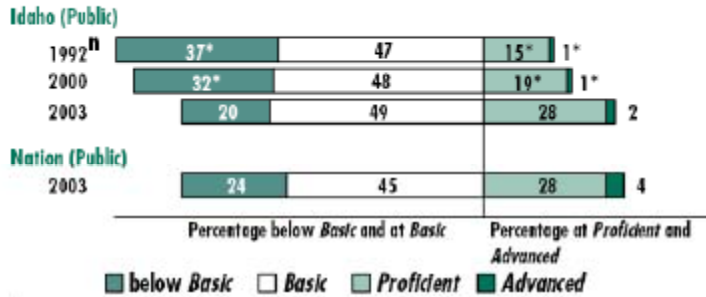
The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American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

**NAEP’s “percent At or Above *Basic*” is the most directly comparable statistic for confirming state AYP results.**

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.

### Student Percentage at NAEP Achievement Levels

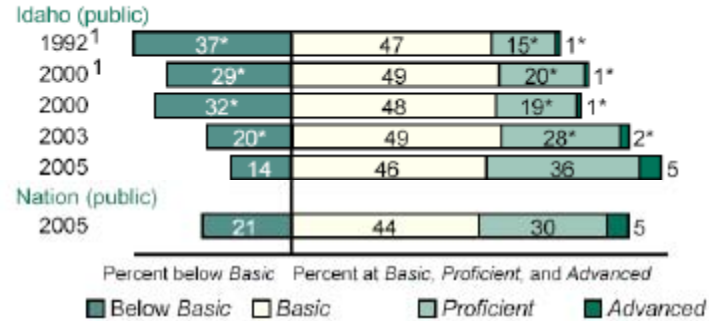


<sup>1</sup> Accommodations were not permitted for this assessment.

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



<sup>1</sup> 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.

**Reading Framework for the  
2007 National Assessment  
of Educational Progress**

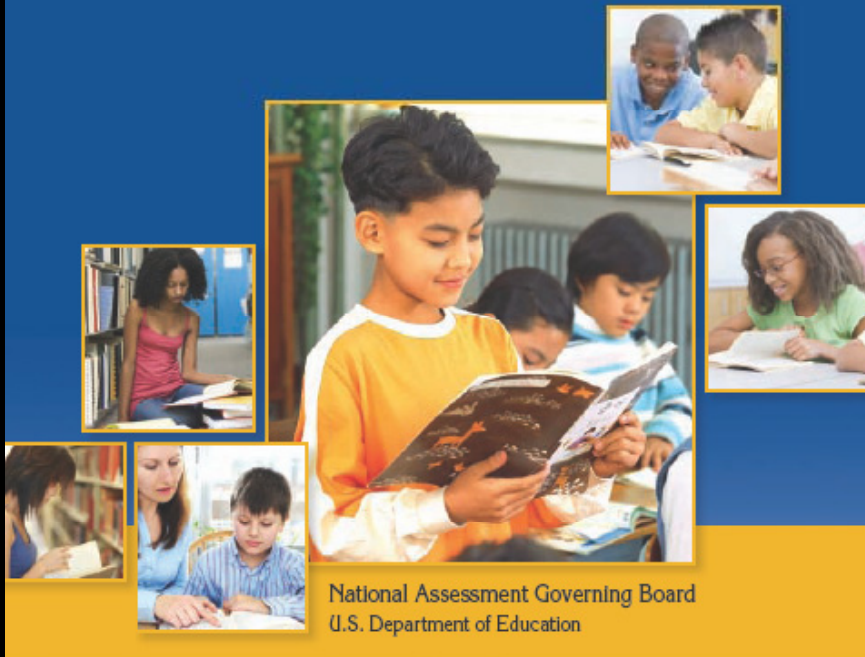


**National Assessment Governing Board  
U.S. Department of Education**

**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 2007**

Reading Framework  
for the 2009  
National Assessment  
of Educational Progress



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

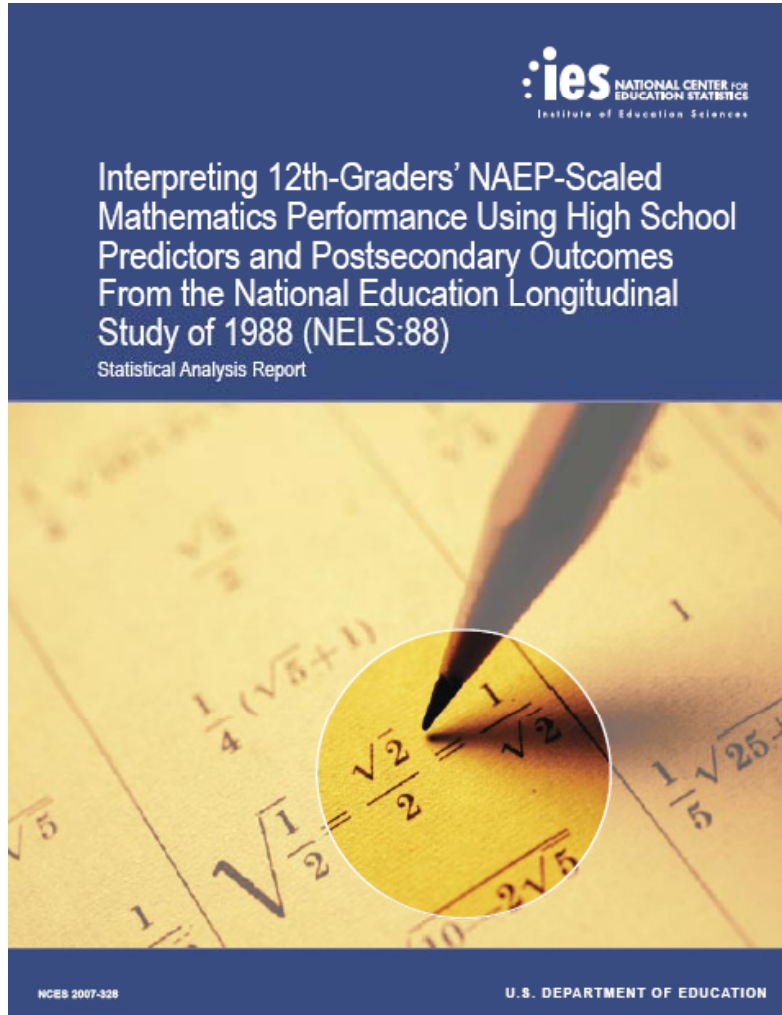
National Assessment Governing Board.  
(2008). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.



## “Letter Grades” for NAEP Achievement Levels

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 of challenging content	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

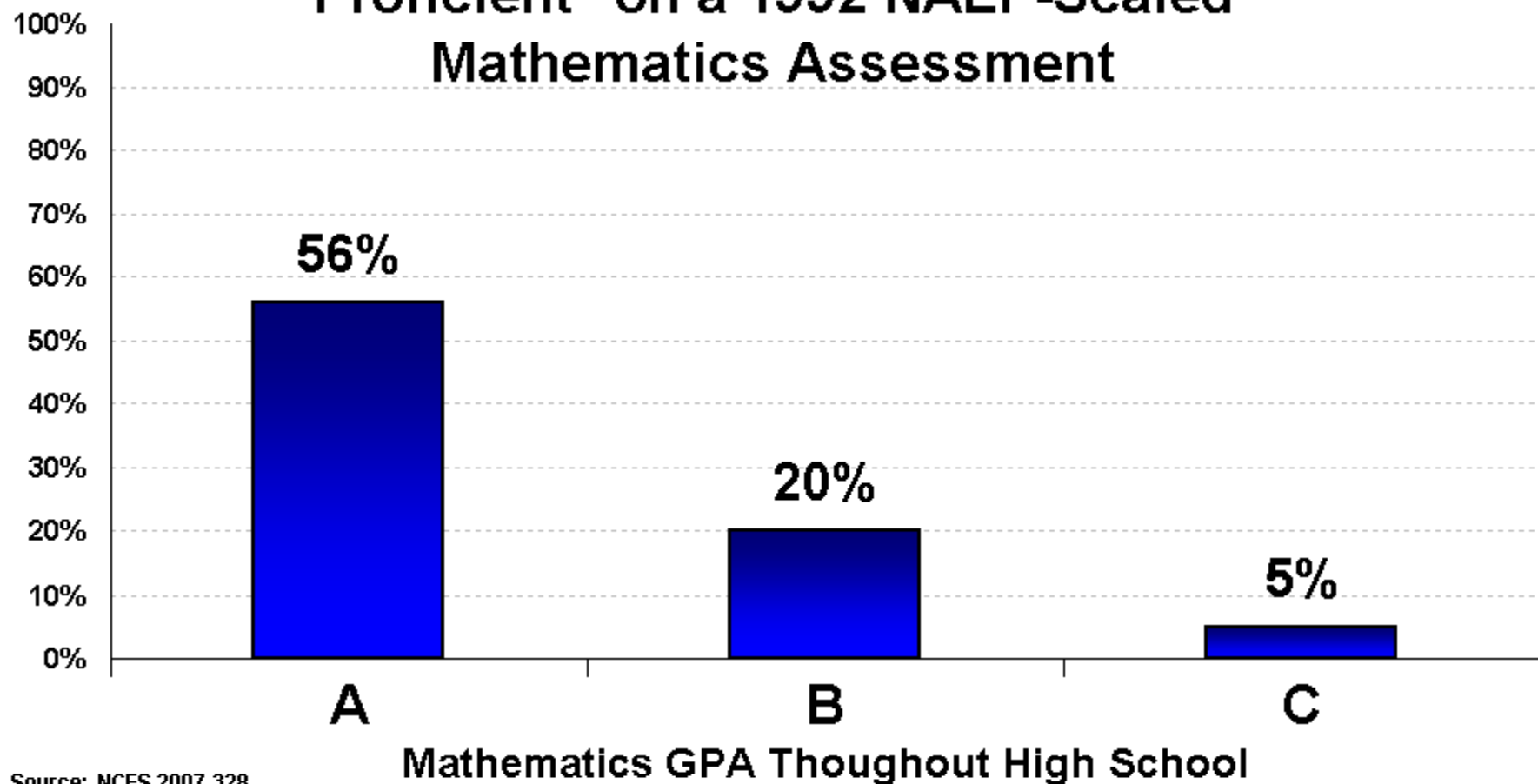
Stoneberg, B.D. (2007). Using NAEP to Confirm State Test Results in the No Child Left Behind Act. *Practical Assessment Research & Evaluation*, 12(5). Available online: <http://www.pareonline.net/pdf/v12n5.pdf>



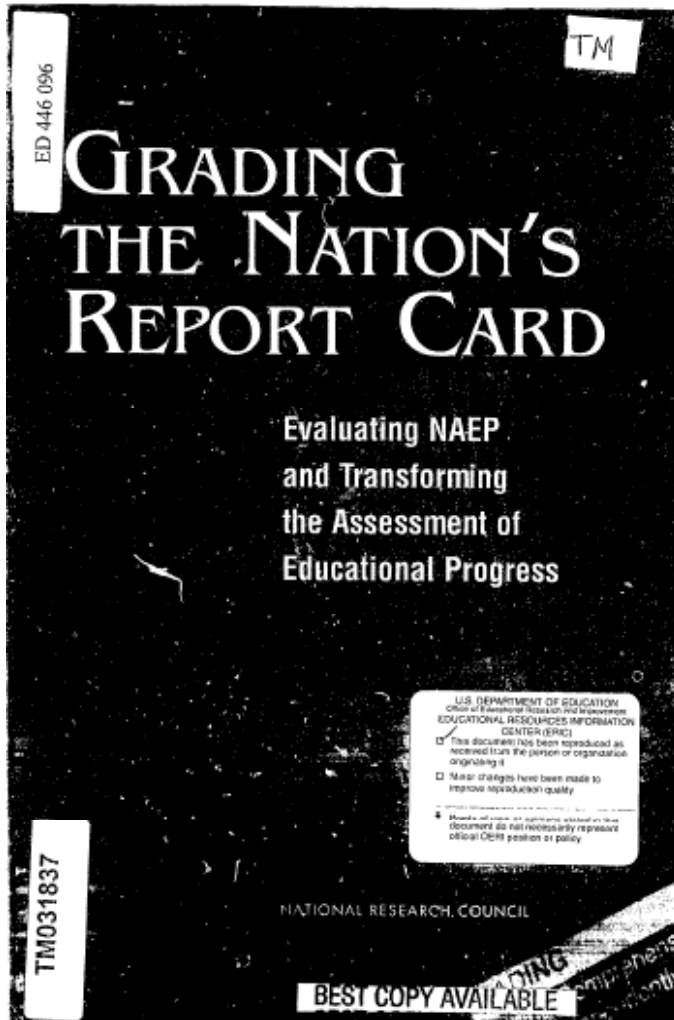
**A majority (56 %) of Proficient and above performers on the 1992 NAEP-scaled mathematics assessment maintained an “A” average in mathematics throughout high school. Some 20 percent of “B” students and 5 percent of “C” students reached the proficient or advanced levels.**

Scott, L.A., and Ingels, S.J. (2007). *Interpreting 12th-graders' NAEP-scaled mathematics performance using high school predictors and postsecondary outcomes from the National Education Longitudinal Study of 1988 (NELS:88)* (NCES 2007-328). Washington, D.C.: National Center for Education Statistics, U.S. Department of Education.

## The Percentage of High School Seniors by Mathematics GPA Who Scored "At or Above Proficient" on a 1992 NAEP-Scaled Mathematics Assessment



Source: NCES 2007-328



**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...the reports should strongly and clearly emphasize that 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...rather than focusing on the percentages in each category in a single year.**

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.

# THE NATION'S REPORT CARD

---

National Assessment of Educational Progress

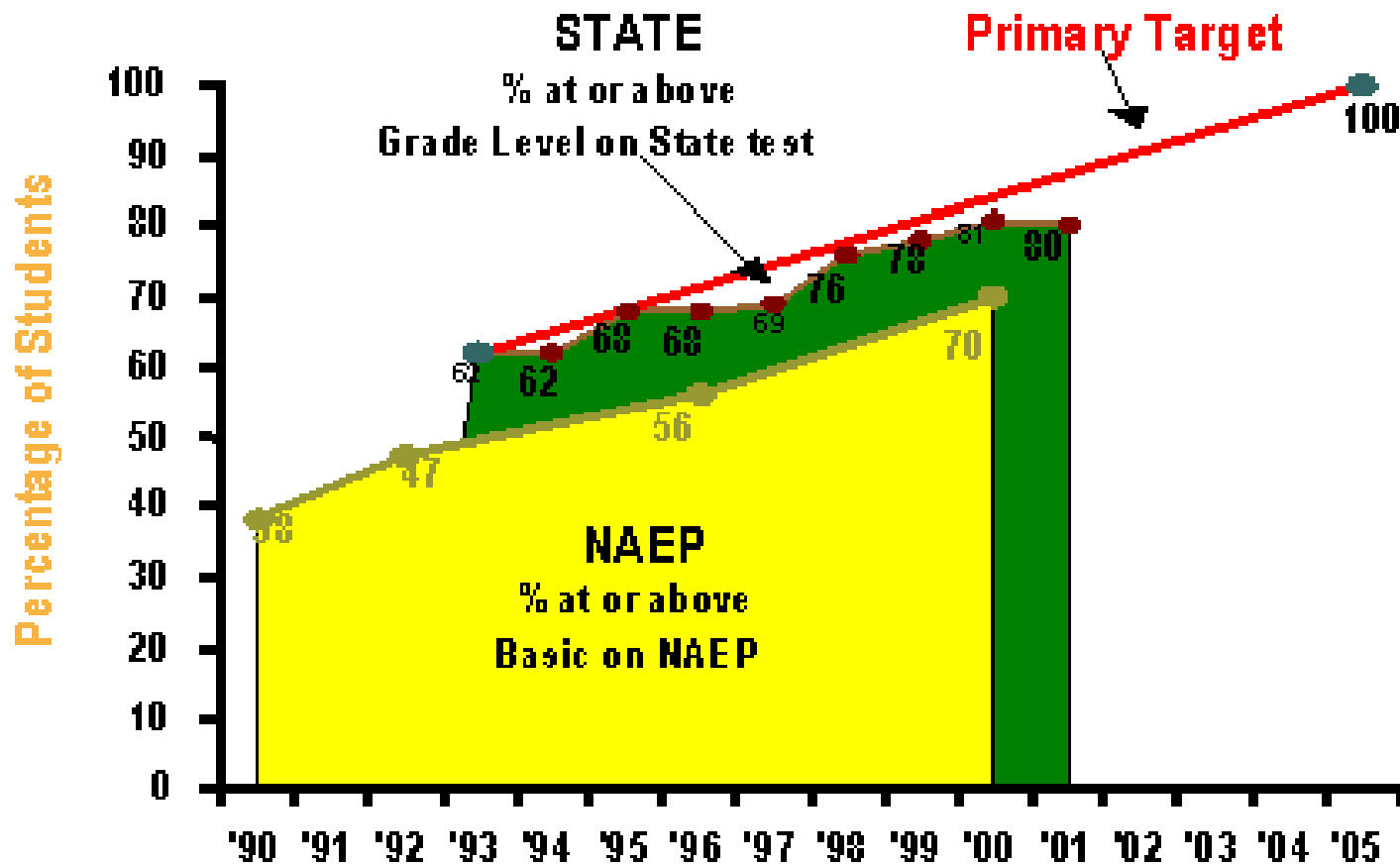
---

**The 2001 reauthorization law requires that the achievement levels be used on a trial basis until the Commissioner of Education Statistics determines that the achievement levels are "reasonable, valid, and informative to the public" (P.L. 107-110, 115 Stat. 1425 [2002]). Until that determination is made, the law requires the Commissioner and the National Assessment Governing Board to state clearly the trial status of the achievement levels in all NAEP reports.**

**A proven alternative to the current process has not yet been identified.**

**The Board and NCES believe that the [achievement levels are useful for reporting trends](#) in the educational achievement of students in the United States. However, [...] NCES concludes that these [achievement levels should continue to be used on a trial basis and should continue to be interpreted and used with caution](#).**

See <http://nces.ed.gov/nationsreportcard/achlevdev.asp?id=re>

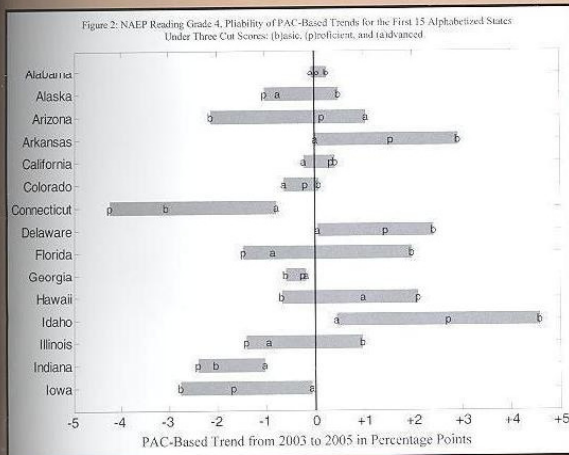


Carr, P.G. (2002, August). *Legislative and Policy Update*. PowerPoint presentation at the National Assessment of Educational Progress (NAEP) State Coordinator Two-day Orientation of the NAEP State Service Center, Washington, D.C.

# EDUCATIONAL MEASUREMENT: ISSUES AND PRACTICE

Volume 26, Number 4

Winter 2007



Reliability as Argument  
*Jay Parkes*

Discrepancies Between Score Trends  
from NAEP and State Tests:  
A Scale-Invariant Perspective  
*Andrew D. Ho*

Subscores Based on Classical Test  
Theory: To Report or Not to Report  
*Sandip Sinharay, Shelby Haberman,  
and Gautam Puhari*

Validity Issues in Test Speededness  
*Ying Lu and Stephen G. Sireci*

An NCME Instructional Module  
on Estimating Item Response  
Theory Models Using Markov  
Chain Monte Carlo Methods  
*Jee-Seon Kim and Daniel M. Boli*

2008 NCME Annual Meeting Ad  
2008 NCME Election Slate

NCME national  
council on  
measurement  
in education

**“Trend comparisons require both technical care and substantive consideration. As useful as PAC [percent above cut-score] statistics have been in communicating test results to the public, their properties as trend statistics render them ill-suited for trend comparison.”**

Ho, A.D. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice*, 26(4), pp. 11-20.

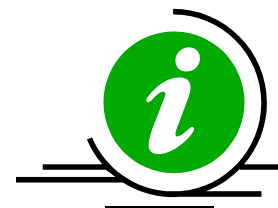
# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal. ISSN 1531-7714

Search:

- 
- 2007 **Using NAEP to Confirm State Test Results in the No Child Left Behind Act** *Stoneberg, Bert D.*  
Viewed 10,192 times since 5/25/2007.
  - 2005 **Please Don't Use NAEP Scores to Rank Order the 50 States** *Stoneberg, Bert D.*  
Viewed 10,072 times since 8/22/2005.

<http://pareonline.net/genpare.asp?wh=4&abt=stoneberg>





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

Stoneberg, B.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) National Conference on Large-Scale Assessment, Nashville, TN. Available online: <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED497395>

**Guest Editorial: Martin Harris on the NAEP**

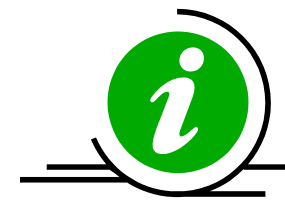
<http://vttradcon.wordpress.com/2008/09/25/guest-editorial-martin-harris-on-the-naep/>

***A Funny Thing Happened on the Way to the Forum (Part I of IV)***

My literary betters [ ] can plagiarize far more skillfully than I, and so my theft of the movie title “A Funny Thing...” to head this column refers not to the Roman Forum but rather to a colloquium of educators in –where else—Nashville [ ]. The official educator forum in the Volunteer State in mid-June of last year went unpublicized [ ] and I knew nothing of it until a few short weeks ago when I contacted the US Department of Education for an explanation of a puzzling subject in public education: the substantial discrepancy in student achievement test scores between the federal NAEP tests and all the State-preferred local tests.

[ ] Until recently, no one paid much attention to this National Assessment of Educational Progress (NAEP) even though the resulting student test scores were uniformly quite dismal, [ ] about 2/3 of all test-takers couldn't make “proficient” [ ] and couldn't, therefore, function at grade level. [ ] Such an intractable problem calls for a conference; or, if you prefer, a Forum.

And then a funny thing happened on the way to the Forum (or maybe once there, Quisnam teneo? Who knows?) a solution to the problem was discovered, or created, or invented. It can be found on pages 8 and 9 of the conference – oops, Forum—report, “An Explanation for the Large Differences Between State and NAEP “Proficiency” Scores Reported for Reading in 2005”.



## **Common (but False) Assumptions**

- ▶ **American public schools are a dismal failure ...**
- ▶ **Proficient is proficient is proficient ...**
- ▶ **A test is a test is a test ...**
- ▶ **Everyone is entitled to his or her own belief about how to interpret and use a test score ...**

<b>Advanced</b>	<b>This level signifies superior performance.</b>
<b>Proficient</b>	<b>This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.</b>
<b>Basic</b>	<b>This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.</b>

**Basic denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work.**

**A grade of C- to B denotes partial mastery of prerequisite knowledge and skills that are fundamental for B+ to A work.**