

NAEP State Service Center Spring Assessment Workshop  
Doubletree Hotel Bethesda -- Bethesda, Maryland  
March 20-22, 2007

## The Valid Use of NAEP Achievement Level Scores to Confirm State Testing Results in the No Child Left Behind Act

Bert D. Stoneberg  
NAEP Program Manager  
Office of the Idaho State Board of Education

### ABSTRACT

The No Child Left Behind Act sanctions the use of NAEP scores to confirm state testing results. The U.S. Department of Education, as test developer, is responsible to set forth how NAEP scores are to be interpreted and used. Thus far, the Department has not published a clear set of guidelines for using NAEP achievement level scores to conduct a confirmation analysis. The lack of an official guidance document, however, does not mean that the issue has been ignored. This study searched the literature from the Department and from other sources to locate principles or “ground rules” that inform and guide confirmation analyses. The basic principles that the study identified do not exhaust all factors relevant to conducting a confirming analysis. Taken together, however, these principles provide researchers ample direction regarding the valid use of NAEP achievement level scores. They include:

- NAEP’s definition of *Proficient* is not synonymous with proficiency in a subject.
- Confirmation should not be conducted on a point-by-point basis.
- Differences between NAEP and the state testing program must be explored and reported.
- NAEP’s percentage at or above *Basic* is the most directly comparable statistic for confirming state AYP results.
- Plotting and comparing “proficiency trend lines” from NAEP and the state test is a defensible method for using NAEP to confirm state results.

(8 references, 1 table, 2 figures, and 1 appendix, a Power Point handout).

NAEP State Service Center Spring Assessment Workshop  
Doubletree Hotel Bethesda -- Bethesda, Maryland  
March 20-22, 2007

## The Valid Use of NAEP Achievement Level Scores to Confirm State Testing Results in the No Child Left Behind Act

Bert D. Stoneberg  
NAEP Program Manager  
Office of the Idaho State Board of Education

Six basic principles address the valid use of achievement level scores from the National Assessment of Educational Progress (NAEP) to confirm state testing results. There are other important “ground rules” in the literature that should also be considered when using NAEP achievement levels, but the time today for this presentation is limited. The basic principals for the discussion today are:

- The U.S. Department of Education (USED) is responsible to set forth how NAEP scores are to be interpreted and used. (1999)
- NAEP’s definition of *Proficient* is not synonymous with grade-level “proficiency in a subject.” (2001)
- Confirmation of state Adequate Yearly Progress (AYP) results should not be conducted on a point-by-point basis. (2002)
- When confirming state AYP results, differences between NAEP and the state testing program must be explored & reported. (2002)
- NAEP’s percentage At or Above *Basic* is the most directly comparable statistic for confirming state AYP results. (2004)
- Comparison of “proficiency trend lines” from NAEP and the state test is a defensible method for using NAEP to confirm state AYP results. (1998, 2002)

**{1} The U.S. Department of Education (USED) is responsible to set forth how NAEP scores are 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 has assigned the role of test developer to the U.S. Department of Education. This being the case, Standard 1.2 places responsibility upon USED to specify the appropriate interpretation and use of NAEP achievement level scores. The lead groups in USED implementing NAEP are the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES). The Office of Elementary and Secondary Education in the U.S. Department of Education deals with NCLB.

NAEP was around more than three decades before the No Child Left Behind Act (NCLB) mandated a new use for the assessment. USED 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 guidelines document from the developer does not constitute license to use NAEP achievement level scores haphazardly or without caution. There is sufficient information about the

topic in publications from USED and from external sources, however, that a few basic principles or “ground rules” may be established.

NAEP releases achievement level summary results 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 use NAEP data that has not been previously validated upon the user whether member of the public (including the media) or educational researcher.

**{2} NAEP’s definition of *Proficient* is not synonymous with grade-level “proficiency in a subject.”**

In 2001, NAGB published a series of booklets to inform the general public about the use and interpretation of NAEP achievement levels. The following text is a section of the reading booklet “How Should Achievement Levels Be Interpreted”, but the identical language is also found in the booklets prepared 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).

NAEP achievement levels do not attend to grade-level performance. 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 the states must attend to proficiency in the 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).

Table 1 represents a personal attempt to further understand USED’s language about the NAEP achievement levels by “assigning” a range of letter grades to the language that one might see on the report cards of students performing at each NAEP achievement level. The letter grades are based on

the author’s thirty plus years of experience in the public schools of Washington, Oregon and Idaho, and on a hazy, general perception about how students seem to be distributed across the achievement levels and letter grades. [Feel free to question the assigned grades or to replace them with your own. ]

**Table 1. English language descriptors for each NAEP achievement level from NAGB’s achievement level report for reading (Loomis & Bourque, 2001b), and the estimated range of “letter grades” for 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 Mastery	A   B+
<i>Basic</i>	Proficiency in subject (common meaning) More than minimal competency	B   C-
<i>Below Basic</i>	Minimally competent	D+   F

**{3} Confirmation of state Adequate Yearly Progress (AYP) results should not be conducted on a point-by-point basis.**

In 2002, NAGB released a report from the Ad Hoc committee it had convened to study how the Secretary of Education might use NAEP achievement levels to confirm state testing results as NCLB permitted (Ad Hoc Committee, 2002). The committee recommended that using NAEP to confirm state test results should not be conducted on a point-by-point basis.

The purpose of using NAEP is to provide a second “snapshot” of state results. The Ad Hoc Committee noted, “A number of factors exist that potentially limit the degree of convergence between NAEP and state test results.” An single object can present quite different images to the eye when color film or black and white film are used. In different light settings, two unlike objects may appear to the eye as being the same object.

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

The Ad Hoc Committee (2002) identified differences between NAEP and the state testing program as factors that might limit the convergence of their results. It noted:

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.

The Ad Hoc Committee’s list identified eight plus “etc.” potential differences that could justify interpreting identical percentage scores from the state and NAEP tests as being quite different. There are, no doubt, more.

**{5} NAEP’s percentage At or Above *Basic* is the most directly comparable statistic for confirming state AYP results.**

In 2004, the NAEP Validity Studies Panel released its finding from a statistical study that the percent at or above *Basic* is the appropriate NAEP statistic to use when confirming state AYP results. The Panel was established by NCES under contract 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. 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 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 for 2005 some of the narratives, tables and charts prominently displayed on 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.

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

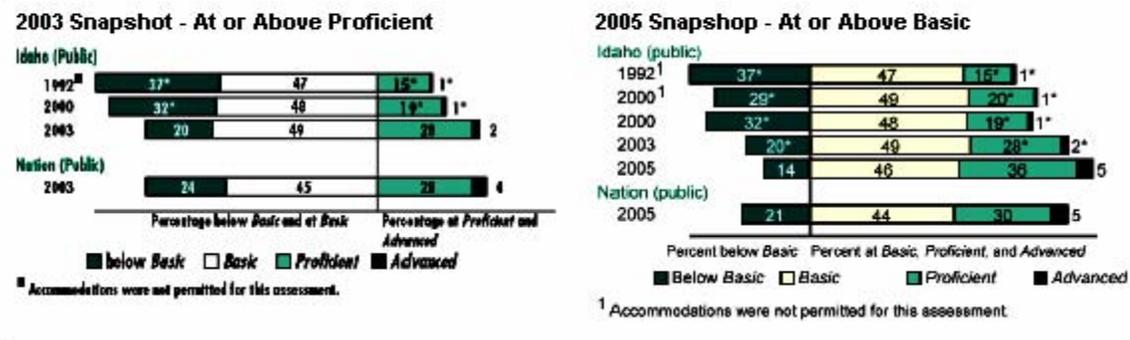


Figure 1. An illustration of changing the reporting focus from Percent At or Above *Proficient* to Percent At or Above *Basic* on the trend charts from the NAEP grade 4 Mathematics Snapshot Report for Idaho between 2003 and 2005.

{6} Comparison of “proficiency trend lines” from NAEP and the state test is a defensible method for using NAEP to confirm state AYP results.

The Ad Hoc Committee for Confirming Test Results -- after posting a long list of cautions and avoids – did recommend one use of NAEP scores to confirm state test results. NAEP achievement levels can be used as evidence to confirm the general trend of state test AYP results in grades 4&8 reading and

mathematics (Ad Hoc Committee, 2002). This use coincides with the original purpose of NAEP, namely to measure 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 negative. The Academy, however, 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).

## Bringing it all together...

Figure 2 illustrates how NAEP might be used to confirm state testing results (Carr, 2002). It's a useful graphic for bringing the discussion points of this paper together. 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 2 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.

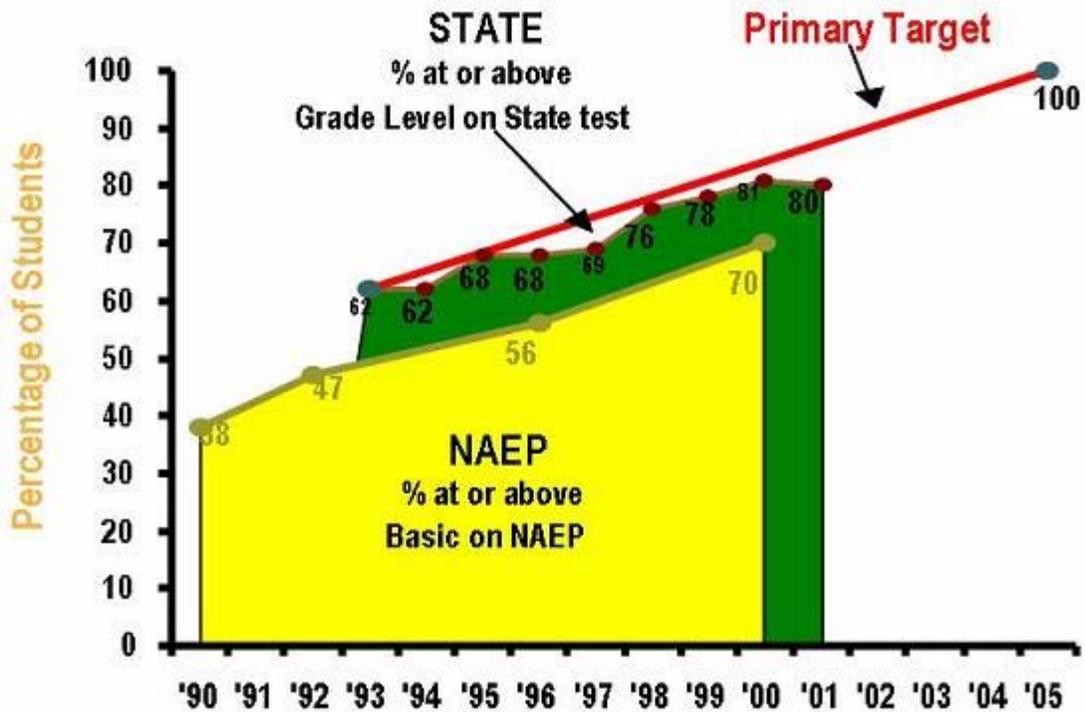


Figure 2. Pre-NCLB graphic (courtesy of Wendy Yen) illustrating how NAEP percent at or above *Basic* might be used to confirm state testing results in the No Child Left Behind Act of 2001.

# # #

## 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](http://www.nagb.org/pubs/color_document.pdf)

Carr, P. (2002, August). *Legislative and Policy Update*. Paper presented 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.: 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](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>

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>

### **Citation:**

Stoneberg, B.D. (2007, March). *The Valid Use of NAEP Achievement Level Scores to Confirm State Testing Results in the No Child Left Behind Act*. Paper presented at the National Assessment of Educational Progress (NAEP) State Service Center Spring Assessment Workshop, Bethesda, MD.

NAEP State Service Center Spring Assessment Workshop  
Doubletree Hotel Bethesda -- Bethesda, Maryland  
March 22, 2007

## **The Valid Use of NAEP Achievement Level Scores to Confirm State Testing Results in the No Child Left Behind Act**



Bert Stoneberg  
NAEP Program Manager  
Office of the State Board of Education  
Boise, Idaho  
bstoneberg@osbe.idaho.gov

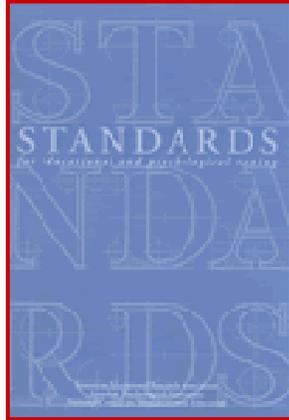
1

## **The Valid Use of NAEP Achievement Level Scores**

1. The U.S. Department of Education is responsible to set forth how NAEP scores are to be interpreted and used. (1999)
2. NAEP's definition of "Proficient" is not synonymous with "proficiency in a subject." (2001)
3. Confirmation of state AYP results should NOT be conducted on a point-by-point basis. (2002)
4. When confirming state AYP results, differences between NAEP and the state testing program must be explored & reported. (2002)
5. NAEP percentage "At or Above Basic" is the most directly comparable statistic for confirming state AYP results. (2004)
6. Comparison of "proficiency" trend lines from NAEP and the state test is a defensible method for using NAEP to confirm state AYP results. (2002, 1998)

2

# 1. 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.

3

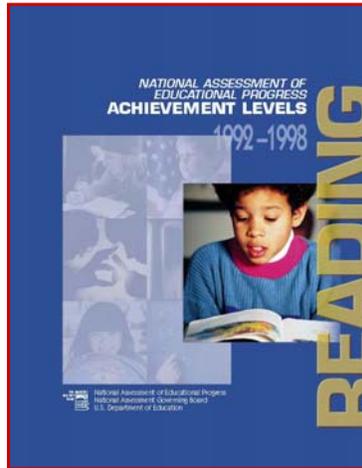
## Valid Use of Test Scores

- ▶ **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.

4

## 2. NAEP's definition of "Proficient" is not synonymous with "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.

5

## How Should Achievement Levels Be Interpreted?

“Unlike most assessments, there are no individual scores on NAEP. 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.”

6

Achievement Level	NAEP Achievement Level Descriptors	Letter Grade Range
Advanced		A+
Proficient	Some of the best students you know	A ↓
	Mastery	B+
Basic	Proficiency in subject (common meaning)	B ↓
	More than minimal competency	C-
Below Basic	Minimally competent	D+ ↓ F

**Descriptors and estimated letter grade ranges for NAEP achievement levels.**

7

**NCLB Program Point of View**

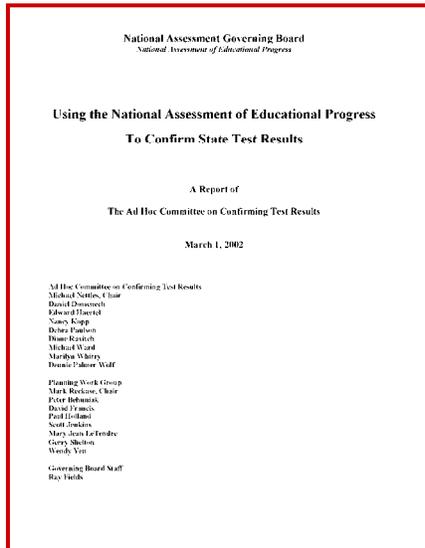
**We remain committed to ensuring that all students can read and do math at grade level or better by 2014. This is the basic purpose and mission of the No Child Left Behind Act.**

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

**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.**

*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.

8



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

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

9

**South Dakota and Idaho both reported 87 percent proficient as their 2005 AYP result for reading in grade 4, 87 = 87?**

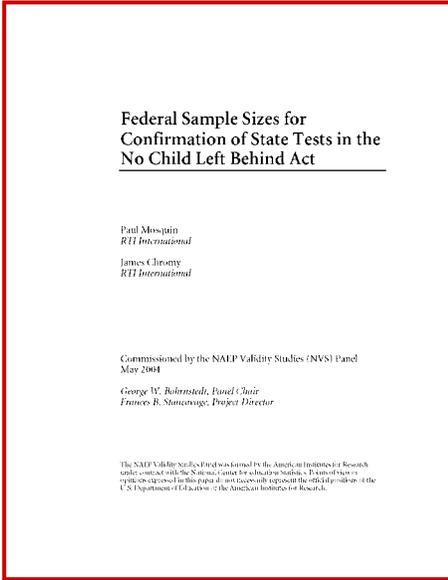
**“Potential differences between NAEP and state testing programs include: content coverage in the subjects, definitions of subgroups, changes in the demography within a state over time, sampling procedures, standard-setting approaches, reporting metrics, student motivation in taking the state test versus 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.

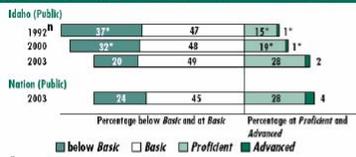
10

**5. 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.

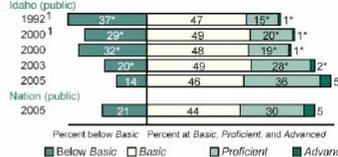


**Student Percentage at NAEP Achievement Levels**



Idaho Snapshot Report  
Mathematics 2003, Gr 4

**Student Percentage at NAEP Achievement Levels**



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 Study Panel's recommendations.**

## 6. Comparison of "proficiency" trend lines from NAEP and the state test is a defensible method for using NAEP to confirm state AYP results.

**NAEP can be used as evidence to confirm the general trend of state test results in grades 4&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 (below basic, basic, proficient, and advanced), 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.

13

## Bringing it all together...

