National Assessment Governing Board
Reporting and Dissemination Committee

December 6, 2013
10:00 a.m.-12:45 p.m.

AGENDA

10:00 – 10:45 am	NAEP Testing and Reporting on Students with Disabilities and English Language Learners *Larry Feinberg, NAGB Staff* *Grady Wilburn, NCES* **[Joint meeting with COSDAM]**	Attachment A
10:45 – 11:15 am	Embargo Policy for National Assessment Reports *NAGB Staff*	Attachment B
11:15 – 11:40 am	Board and Committee Input into NAEP Reports *Stephaan Harris, NAGB Staff*	Attachment C
11:40 – 11:50 am	Review of NAEP 2013 Reading and Mathematics Release *Stephaan Harris, NAGB Staff* *Amy Buckley, Reingold*	Attachment D
11:50 – 11:55 am	**ACTION:** Release Plan for 2013 TUDA Report Cards *Stephaan Harris, NAGB Staff*	Attachment E
11:55 am – 12:10 pm	Board Chairman's Charge to the Committee for 2014 *David Driscoll, NAGB Chair*	
12:10 pm – 12:45 pm	NAEP Contextual Variables a. **ACTION:** Contextual Information Framework b. Using NAEP for Education Indicators *Alan Ginsburg and Marshall Smith, Consultants* **[Joint Meeting with ADC]**	Attachment F Attachment G
Information Items	• Projected Schedule for Future NAEP Reports • NAGB Parent Summit	Attachment H Attachment I

NAEP Testing and Reporting on Students with Disabilities and English Language Learners

The inclusion of students with disabilities (SD) and English language learners (ELL) is key to the success of the Nation's Report Card. It is important that we measure all of the students in NAEP jurisdictions to give the most accurate picture of student knowledge and skills in those areas. The National Assessment Governing Board, believing that the validity of NAEP results has been threatened by high exclusion rates and substantial variations from state-to-state, has adopted a policy to promote increased inclusion and testing of these two important groups.

Since the policy was enacted in March 2010, we have seen inclusion rates improve across all subjects and grades. For example, ten years ago in 4th grade reading 19 states would not have met the 95 percent inclusion goal that the Board policy set. In 2013, that number dropped to one state. In fact, in 2013 only one state—Maryland in both 4th and 8th grade reading—did not meet the 95 percent inclusion goal across all grades/subjects. The exclusion rates in 2013 are the lowest since state NAEP began more than 20 years ago.

The December presentation to the joint committees of COSDAM and Reporting and Dissemination will focus on (1) improvements resulting in the lowest exclusion rates in NAEP's history, and (2) NCES plans for implementing the follow-up actions endorsed at the Governing Board meeting in May 2013. These include the following:

- giving greater prominence to full population estimates, which present adjusted scores that seek to account for differences in students excluded,
- providing additional information online on the proportion of students excluded because they use an accommodation NAEP does not allow, and
- developing a special report after the main data release that includes a full discussion of exclusion, participation and refusal issues, and the total participation rate for each jurisdiction in NAEP.

The following materials are attached:

A.1 Governing Board Policy on NAEP Testing and Reporting on Students with Disabilities and English Language Learners – Adopted March 6, 2010

A.2 Action on reporting SD and ELL data as part of 2013 NAEP results—Joint meeting of Committee on Standards, Design and Methodology and Reporting and Dissemination—May 17, 2013

A.3 State-by-state data on exclusions, including proportion of students excluded because they use an accommodation on state tests that is not permitted by NAEP [Available in Excel]

A.4 NAEP Accommodations—Permitted and Used (2013)

A.5 Read-aloud accommodation

A.6 Calculator accommodation

A.7 Full-population estimates

A.8 NAEP 2013 data on students assessed, accommodated, excluded, and absent, including total participation rate [Available in PDF and Excel]

A.9 Former English-language learners

ADOPTED UNANIMOUSLY BY NAGB—3/6/2010

National Assessment Governing Board

Policy Statement on NAEP Testing and Reporting on Students with Disabilities and English Language Learners

INTRODUCTION

To serve as the Nation's Report Card, the National Assessment of Educational Progress (NAEP) must produce valid, comparable data on the academic achievement of American students. Public confidence in NAEP results must be high. But in recent years it has been threatened by continuing, substantial variations in exclusion rates for students with disabilities (SD) and English language learners (ELL) among the states and urban districts taking part.

Student participation in NAEP is voluntary, and the assessment is prohibited by law from providing results for individual children or schools. But NAEP's national, state, and district results are closely scrutinized, and the National Assessment Governing Board (NAGB) believes NAEP must act affirmatively to ensure that the samples reported are truly representative and that public confidence is maintained.

To ensure that NAEP is fully representative, a very high proportion of the students selected must participate in its samples, including students with disabilities and English language learners. Exclusion of such students must be minimized; they should be counted in the Nation's Report Card. Accommodations should be offered to make the assessment accessible, but these changes from standard test administration procedures should not alter the knowledge and skills being assessed.

The following policies and guidelines are based on recommendations by expert panels convened by the Governing Board to propose uniform national rules for NAEP testing of SD and ELL students. The Board has also taken into consideration the views expressed in a wide range of public comment and in detailed analyses provided by the National Center for Education Statistics, which is responsible for conducting the assessment under the policy guidance of the Board**.** The policies are presented not as statistically-derived standards but as policy guidelines intended to maximize student participation, minimize the potential for bias, promote fair comparisons, and maintain trends. They signify the Board's strong belief that NAEP must retain public confidence that it is fair and fully-representative of the jurisdictions and groups on which the assessment reports.

POLICY PRINCIPLES

1. As many students as possible should be encouraged to participate in the National Assessment. Accommodations should be offered, if necessary, to enable students with disabilities and English language learners to participate, but should not alter the constructs assessed, as defined in assessment frameworks approved by the National Assessment Governing Board.

2. To attain comparable inclusion rates across states and districts, special efforts should be made to inform and solicit the cooperation of state and local officials, including school personnel who decide upon the participation of individual students.

3. The proportion of all students excluded from any NAEP sample should not exceed 5 percent. Samples falling below this goal shall be prominently designated in reports as not attaining the desired inclusion rate of 95 percent.

4. Among students classified as either ELL or SD a goal of 85 percent inclusion shall be established. National, state, and district samples falling below this goal shall be identified in NAEP reporting.

5. In assessment frameworks adopted by the Board, the constructs to be tested should be carefully defined, and allowable accommodations should be identified.

6. All items and directions in NAEP assessments should be clearly written and free of linguistic complexity irrelevant to the constructs assessed.

7. Enhanced efforts should be made to provide a short clear description of the purpose and value of NAEP and of full student participation in the assessment. These materials should be aimed at school personnel, state officials, and the general public, including the parents of students with disabilities and English language learners. The materials should emphasize that NAEP provides important information on academic progress and that all groups of students should be counted in the Nation's Report Card. The materials should state clearly that NAEP gives no results for individual students or schools, and can have no impact on student status, grades, or placement decisions.

8. Before each state and district-level assessment NAEP program representatives should meet with testing directors and officials concerned with SD and ELL students to explain NAEP inclusion rules. The concerns of state and local decision makers should be discussed.

IMPLEMENTATION GUIDELINES

For Students with Disabilities

1. Students with disabilities should participate in the National Assessment with or without allowable accommodations, as needed. Allowable accommodations are any changes from standard test administration procedures, needed to provide fair access by students with disabilities that do not alter the constructs being measured and produce valid results. In cases where non-standard procedures are permitted on state tests but not allowed on NAEP, students will be urged to take NAEP without them, but these students may use other allowable accommodations that they need.

2. The decision tree for participation of students with disabilities in NAEP shall be as follows:

NAEP Decision Tree for Students with Disabilities

BACKGROUND CONTEXT

1. NAEP is designed to measure constructs carefully defined in assessment frameworks adopted by the National Assessment Governing Board.

2. NAEP provides a list of appropriate accommodations and non-allowed modifications in each subject. An appropriate accommodation changes the way NAEP is normally administered to enable a student to take the test but does not alter the construct being measured. An inappropriate modification changes the way NAEP is normally administered but does alter the construct being measured.

STEPS OF THE DECISION TREE

3. In deciding how a student will participate in NAEP:

   a. If the student has an Individualized Education Program (IEP) or Section 504 plan and is tested without accommodation, then he or she takes NAEP without accommodation.

   b. If the student's IEP or 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.

   c. If the student's IEP or 504 plan specifies an accommodation or modification that is not allowed on NAEP, then the student is encouraged to take NAEP without that accommodation or modification.

3. Students should be considered for exclusion from NAEP only if they have previously been identified in an Individualized Education Program (IEP) as having the most significant cognitive disabilities, and are assessed by the state on an alternate assessment based on alternate achievement standards (AA-AAS). All students tested by the state on an alternate assessment with modified achievement standards (AA-MAS) should be included in the National Assessment.

4. Students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.

5. NAEP should report separately on students with Individualized Education Programs (IEPs) and those with Section 504 plans, but (except to maintain trend) should only count the students with IEPs as students with disabilities. All 504 students should participate in NAEP.

   At present the National Assessment reports on students with disabilities by combining results for those with an individualized education program (who receive special education services under the Individuals with Disabilities Education Act [IDEA]) and students with Section 504 plans under the Rehabilitation Act of 1973 (a much smaller group with disabilities who are not receiving services under IDEA but may be allowed test accommodations).[*] Under the Elementary and Secondary Education Act, only those with an IEP are counted as students with disabilities in reporting state test results. NAEP should be consistent with this practice. However, to preserve trend, results for both categories should be combined for several more assessment years, but over time NAEP should report as students with disabilities only those who have an IEP.

6. Only students with an IEP or Section 504 plan are eligible for accommodations on NAEP. States are urged to adopt policies providing that such documents should address participation in the National Assessment.

For English Language Learners

1. All English language learners selected for the NAEP sample who have been in United States schools for one year or more should be included in the National Assessment. Those in U.S. schools for less than one year should take the assessment if it is available in the student's primary language.

   One year or more shall be defined as one full academic year before the year of the assessment.

[*] NOTE: The regulation implementing Section 504 defines a person with a disability as one who has a physical or mental impairment which substantially limits one or more major life activities, has a record of such an impairment, or is regarded as having such an impairment. 34 C.F.R. § 104.3(j)(1).

2. Accommodations should be offered that maximize meaningful participation, are responsive to the student's level of English proficiency, and maintain the constructs in the NAEP framework. A list of allowable accommodations should be prepared by NAEP and furnished to participating schools. Such accommodations may be provided only to students who are not native speakers of English and are currently classified by their schools as English language learners or limited English proficient (LEP).

3. Bilingual versions of NAEP in Spanish and English should be prepared in all subjects, other than reading and writing, to the extent deemed feasible by the National Center for Education Statistics. The assessments of reading and writing should continue to be in English only, as provided for in the NAEP frameworks for these subjects.

4. Staff at each school should select from among appropriate ELL-responsive accommodations allowed by NAEP, including bilingual booklets, those that best meet the linguistic needs of each student. Decisions should be made by a qualified professional familiar with the student, using objective indicators of English proficiency (such as the English language proficiency assessments [ELPA] required by federal law), in accordance with guidance provided by NAEP and subject to review by the NAEP assessment coordinator.

5. Schools may provide word-to-word bilingual dictionaries (without definitions) between English and the student's primary language, except for NAEP reading and writing, which are assessments in English only.

6. NAEP results for ELL students should be disaggregated and reported by detailed information on students' level of English language proficiency, using the best available standardized assessment data. As soon as possible, NAEP should develop its own brief test of English language proficiency to bring consistency to reporting nationwide.

7. Data should be collected, disaggregated, and reported for former English language learners who have been reclassified as English proficient and exited from the ELL category. This should include data on the number of years since students exited ELL services or were reclassified.

8. English language learners who are also classified as students with disabilities should first be given linguistically-appropriate accommodations before determining which additional accommodations may be needed to address any disabilities they may have.

RESEARCH AND DEVELOPMENT

The Governing Board supports an aggressive schedule of research and development in the following areas:

1. The use of plain language and the principles of universal design, including a plain language review of new test items consistent with adopted frameworks.

2. Adaptive testing, either computer-based or paper-and-pencil. Such testing should provide more precise and accurate information than is available at present on low-performing and high-performing groups of students, and may include items appropriate for ELLs at low or intermediate levels of English proficiency. Data produced by such targeted testing should be placed on the common NAEP scale. Students assessed under any new procedures should be able to demonstrate fully their knowledge and skills on a range of material specified in NAEP frameworks.

3. A brief, easily-administered test of English language proficiency to be used for determining whether students should receive a translation, adaptive testing, or other accommodations because of limited English proficiency.

4. The validity and impact of commonly used testing accommodations, such as extended time and small group administration.

5. The identification, measurement, and reporting on academic achievement of students with the most significant cognitive disabilities. This should be done in order to make recommendations on how such students could be included in NAEP in the future.

6. A study of outlier states and districts with notably high or low exclusion rates for either SD or ELL students to identify the characteristics of state policies, the approach of decision makers, and other criteria associated with different inclusion levels.

The Governing Board requests NCES to prepare a research agenda on the topics above. A status report on this research should be presented at the November 2010 meeting of the Board.

National Assessment Governing Board

Committee on Standards, Design and Methodology

May 17, 2013
EXCERPT

JOINT MEETING WITH REPORTING AND DISSEMINATION COMMITTEE

NAEP Testing and Reporting on Students with Disabilities and English Language Learners

The Committees considered two sets of issues: (1) implementation of the Board policy that deals with testing of English language learners (ELL), and (2) reporting options for exclusions, participation rates, and the adjusted scores (termed full-population estimates), which try to take into account the differences in exclusion rates between the states and districts participating in NAEP.

Grady Wilburn, of NCES, gave an update to the committees on these two topics. He noted that under the policy adopted in 2010, the only ELL students that schools may exclude from NAEP are those who have been in U.S. schools for less than one academic year. Even students in this category should not be excluded if NAEP offers a translation in their home language. Students who speak Spanish now account for about 80 percent of ELLs nationwide. NAEP offers Spanish translations of all its tests (in bilingual booklets) except for reading and writing, which under the frameworks adopted by the Board are reading and writing in English.

Mr. Wilburn said the decision tree incorporating the Board policy on which ELL students to test and how to test them was implemented smoothly in the 2013 NAEP. He said exclusion rates went down.

On the reporting issues Mr. Wilburn noted that a joint meeting of the two committees in March had received a full report on implementation of the policy on SD students. Under this policy the only students that may be excluded from NAEP by school personnel are those with the most significant cognitive disabilities who take alternate state assessments with alternate achievement standards, expected to be about 1 percent of enrollment. For practical reasons NCES decided that schools could also continue to exclude students with an individualized education program (IEP) or 504 plan that provides for accommodations on state tests that NAEP does not allow. The non-allowable accommodations in nearly all cases have been read-aloud on the NAEP reading assessment or calculator use on all sections of NAEP math.

In 2013 for the first time NCES permitted students with an IEP requiring calculator use to take calculator-active blocks on NAEP even if they would have been assigned non-calculator blocks as part of the normal NAEP sampling. Mr. Wilburn said a study in 2011 indicated that this would have little impact on results, and, in any case, the number of students involved is small.

By law, student participation in NAEP is voluntary. Parents can refuse to have their children participate for any reason. Under the Board policy, "students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures."

NCES has said doing this would break trends, depress reported scores, and contravene sound psychometric procedures. Under long-standing practice, excluded students are omitted from any calculations of NAEP results, and have no effect on state or district averages. Adjustments are made for refusals or absent students (a much larger group) by reweighting the scores of those with similar characteristics, which tends to lower state and district averages.

There is another analytic procedure, called full-population estimates (FPE), which NCES has used for about a decade to adjust state and district results by imputing scores for excluded SD and ELL students based on the performance of similar SD and ELL students who are tested. Data showing year-to-year changes in the full-population estimates are published on the NAEP website for participating states and districts, but these are given little prominence and do not include the adjusted scores themselves. The FPE scores were provided to the Board at this meeting. They showed most state averages to be about 3 to 6 points lower than reported. In only a few cases were year-to-year changes significantly different.

George Bohrnstedt, of AIR, chair of the NAEP Validity Studies Panel, said his group is concluding a study which shows that FPEs provide less biased results than the current NAEP analysis method, which overstates true scores considerably more.

Another proposal considered for reporting is to publish a total participation rate, based on all students in a sample divided into those tested and not tested for any reason whether excluded, absent, or refused. At present the reported participation rates are calculated after excluded students are subtracted from the number in the sample.

After considerable discussion, the Committees endorsed the following in regard to the reporting of 2013 NAEP results:

1. **Continue previous analysis procedures for exclusions and refusals.**

2. **Give greater prominence and easier accessibility to full-population estimates as part of the information available online at the time of data release.**

3. **Provide additional information online on the proportion of students excluded because they use an accommodation on state tests that is not allowed on NAEP.**

4. **Issue a special report after the main data release with a full discussion of exclusion, participation, and refusal issues that includes data on the total participation rate for each jurisdiction in NAEP.**

Attachment A3 - Excel file: State-by-state data on exclusions, including proportion of students excluded because they use an accommodation on state tests that is not permitted by NAEP

Percentage of fourth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) assessed in NAEP mathematics with accommodations, by SD/ELL category and type of accommodation: 2013

Type of accommodation	SD and/or ELL	SD	ELL
Bilingual dictionary	0.8	0.1	0.8
Braille presentation	#	#	#
Braille response	#	#	#
Breaks	4.1	3.4	1.1
Calculator	1.1	1.1	0.1
Cue to stay on task	1.2	1.1	0.2
Directions read aloud in English	3.3	2.5	1.2
Directions read aloud in Spanish	0.1	#	0.1
Extended time	11.4	8.1	4.4
Large-print booklet	#	#	#
Magnification device	#	#	#
One-on-one	0.6	0.6	0.1
Read aloud (all)	5.7	5.0	1.5
Read aloud (occasional)	1.3	0.8	0.6
Read aloud in Spanish	0.1	#	0.1
School staff administers	0.5	0.5	0.1
Scribe	0.5	0.5	#
Sign language presentation	#	#	#
Sign language response	#	#	#
Small group	9.8	8.0	2.9
Spanish-English booklet	0.3	0.1	0.3
Special equipment	0.4	0.4	#
Other	0.3	0.3	#

Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics Assessment.

Percentage of eighth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) assessed in NAEP mathematics with accommodations, by SD/ELL category and type of accommodation: 2013

Type of accommodation	SD and/or ELL	SD	ELL
Bilingual dictionary	0.7	0.1	0.7
Braille presentation	#	#	#
Braille response	#	#	#
Breaks	2.5	2.3	0.4
Calculator	2.3	2.3	0.2
Cue to stay on task	0.7	0.7	0.1
Directions read aloud in English	2.9	2.5	0.6
Directions read aloud in Spanish	#	#	#
Extended time	9.8	8.2	2.3
Large-print booklet	#	#	#
Magnification device	#	#	#
One-on-one	0.3	0.3	0.1
Read aloud (all)	3.6	3.3	0.5
Read aloud (occasional)	1.3	1.1	0.3
Read aloud in Spanish	#	#	#
School staff administers	0.3	0.3	#
Scribe	0.2	0.2	#
Sign language presentation	#	#	#
Sign language response	#	#	#
Small group	8.4	7.6	1.5
Spanish-English booklet	0.2	#	0.2
Special equipment	0.3	0.3	#
Other	0.3	0.3	#

Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics Assessment.

NAEP Accommodations Increase Inclusiveness

The responses of students with disabilities (SD) and English language learners (ELL) on NAEP assessments represent those of hundreds of other similar students. Without them, information about how to best meet the educational needs of these students would be lost. NAEP incorporates inclusive policies and practices into every aspect of the assessment, including selection of students, participation in the assessment administration, and valid and effective accommodations. This is essential to ensuring an assessment that yields meaningful NAEP results for all students. By representing their peers across the nation on NAEP, students with disabilities and English language learners help to ensure that NAEP results can be used to inform efforts to improve educational programs.

Just like any other student, SD and ELL students are selected to participate in NAEP. Within each selected school and grade to be assessed, students are chosen at random to participate in NAEP. Regardless of race/ethnicity, socioeconomic status, disability, status as an English language learner, or any other factors, every student has the same chance of being selected, because NAEP is administered to a sample of students who represent the student population of the nation as a whole, and for state level tests, of each individual state.

The accommodations allowed on NAEP and those allowed in states are often similar, but there may be some differences. Sometimes these differences result from the way that the subject being measured is defined in the NAEP frameworks. For example, NAEP does not allow read-aloud of any part of the NAEP reading test except the instructions, because decoding words is part of what the NAEP framework is measuring.

The many accommodations available in NAEP can be categorized for ease of understanding:

- Some are regarded as Standard NAEP Practice, available in almost all NAEP assessments for SD and ELL students.
- Other accommodations for SD students require special preparation, such as Braille or signing.
- Other accommodations for ELL students
- Some accommodations are actually built-in features of the computer-based assessments that are available to all students and so are referred to as Universal Design Elements.

Standard NAEP Practice, for SD and/or ELL

Standard Accommodations for SD/ELL Students	*Math*	*Reading*	*Science*	*Writing (CBA)*	*Civics, Economics, Geography, U.S. History*	*Music and Visual Arts*	*TEL (pilot 2013)*	*Writing (before CBA; paper and pencil)*
Extended time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small group, or one-on-one	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
One-on-one	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Directions only read aloud in English	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test items read aloud in English – occasional or most/all	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Breaks during test	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Writes directly in the booklet	No	No	No	No	No	No	No	No

Other Accommodations for SD students

Other Accommodations for SD Students	*Math*	*Reading*	*Science*	*Writing (CBA)*	*Civics, Economics, Geography, U.S. History*	*Music and Visual Arts*	*TEL (pilot 2013)*	*Writing (before CBA; paper and*

								pencil)
Calculator version of the test FN3	Yes FN3	No	No	No	No	No	No	No
Must have an aide present in the testing room	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Responds orally to a scribe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Large print version of the test	Yes	Yes	Yes	Yes UD	Yes	Yes (Music Only)	Yes UD	Yes
Magnification	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Uses template/special equipment/preferential seating	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cueing to stay on task	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Presentation in Braille	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Response in Braille	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Presentation in Sign Language	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Response in Sign Language	Yes	Yes	Yes	Yes	Yes	No	No	No

Other Accommodations for ELL students

Accommodations for ELL Students	***Math***	***Reading***	***Science***	***Writing (CBA)***	***Civics, Economics, Geography, U.S. History***	***Music and Visual Arts***	***TEL (pilot 2013)***	***Writing (before CBA; paper and pencil)***
Bilingual dictionary without definitions in any language	Yes	No	Yes	No	Yes	Yes	Yes	No
Directions only read aloud in Spanish	Yes	Yes	Yes	Yes	Yes	Yes (in Music, No in VisArt)	No	Yes
Spanish/English version of the test (not g12)	Yes	No	Yes	No FN6	Yes FN3	Yes	No	No
Test items read aloud in Spanish (not g12 Math)	Yes	No	Yes	No	Yes	No	No	No

Universal Design Elements and descriptions

Universal Design Elements available for all students in Computer-Based Assessments	**Used for CBA Writing 2011 and TEL pilot 2013**
Adjusting font size	Text size options are provided for short standalone items (e.g., multiple choice and short constructed response items), but not for the scenario-based tasks which will be the standard 14-point font. For the short standalone items, students can change between three sizes ranging from approximately 14-point font to approximately 48-point font. Forty-eight point font is approximately ¾ - inch tall. NOTE: This only enlarges the short standalone items , NOT the scenario-based tasks, tool icons, menus, etc.
Small group	Unnecessary due to the mode of test administration. All students are interacting on a one-on-one basis with the computer and will have earbuds to reduce distractions. Read aloud and other accommodations will be provided through the compute
One-on-one	Unnecessary due to the mode of test administration. All students are interacting on a one-on-one basis with the computer and will have earbuds to reduce distractions. Read aloud and other accommodations will be provided

EDUCATION WEEK
Published Online: October 28, 2013
Published in Print: October 30, 2013, as **'Read-Aloud' Assistance Puts Test-Makers on Spot**

'Read-Aloud' Assistance on Common Tests Proves Contentious

Accommodation for students with print-related disabilities has assessment consortia taking differing approaches

By Christina A. Samuels

Back to Story

Faced with the decision of whether to allow students with dyslexia and other print disabilities the option of having text passages on the common-core tests read aloud to them, the two federally financed consortia responsible for creating the general assessments took a Solomonic approach.

Rather than prohibit the so-called "read-aloud accommodation" entirely or allow reading aloud with no restriction, the Partnership for Assessment of Readiness for College and Careers decided to permit text passages to be read to students, with a notation on score reports saying no claims can be made regarding the student's foundational reading skills. The Smarter Balanced Assessment Consortium opted against the read-aloud accommodation for students in grades 3-5, saying it would invalidate the language constructs being measured; students taking the test in higher grades may use that accommodation.

Like many decisions that attempt to strike a balance between two opposing philosophies—those who wish to leave the use of read-aloud accommodations to a school-based team and those who want to prohibit its use entirely—the decisions of the testing consortia appear to have left no one entirely happy.

And now critics on either side say the consortia's decisions on reading aloud could be setting districts up for violations of special education law, or could ultimately leave a swath of students unable to read because the read-aloud accommodation was used as a crutch.

Fierce Defenders

Richard Allington, a professor of education at the University of Tennessee and one of the country's most recognized experts on early literacy, calls the accommodation "cheating."

"What special education does best is create illiterates," Mr. Allington said. "I know why they don't want their kids tested on reading activity. It's because they've done a terrible job of providing those kids with high-quality reading instruction."

But special educators believe just as strongly that for some children, a read-aloud accommodation is the tool they need to demonstrate what they know.

Moving Beyond the Mainstream

A Common-Core Challenge: Learners With Special Needs

ESL and Classroom Teachers Team Up to Teach Common Core

Common Core Ratchets Up Language Demands for English-Learners

'Read-Aloud' Assistance on Common Tests Proves Contentious

Common Core's Promise Collides With IEP Realities

Tech Assistance in Testing Poses Practical Issues

Common Core Needs Tailoring for Gifted Learners, Advocates Say

Lindsay Jones, the director of public policy and advocacy for the National Center for Learning Disabilities, said she was "stunned" that Smarter Balanced chose to ban read-aloud accommodations for elementary students, even for visually impaired children who may be in the early stages of learning Braille. (An official with the consortium explained that the testing group was told by its legal experts that it could not open up an accommodation to blind children without opening it up to all students with disabilities.)

The learning-disabilities center has argued on behalf of restricting read-aloud accommodations only on test items that gauge print decoding and fluency. Questions that measure other literacy skills, such as picking out a main idea of a text passage, should be open to the accommodation.

"There's been a lot of conjecture and anecdotes about evidence that states overuse read-aloud. I don't know why we would let that dominate our conversation going forward," Ms. Jones said. "Why can't we create items that will test comprehension, that will test decoding? That is a failure, to me, of the promise of these exams."

The restrictions also could represent a conflict with the Individuals with Disabilities Education Act, which gives broad powers to school-based teams of teachers, administrators, parents, and others who have deeper knowledge of a child's capabilities to determine what accommodations may be necessary. Perry A. Zirkel, a professor of education and law at Lehigh University in Bethlehem, Pa., said that the approach taken by Smarter Balanced could be more legally sound than drawing attention to an accommodation. The Educational Testing Service, which administers the SATs, was sued by a student for its policy of flagging tests that were taken under an extended-time accommodation. The testing company agreed to end flagging in 2001.

"My prediction is that [PARCC] thinks it's getting out of the problem," Mr. Zirkel said. "If this gets challenged, they will likely follow the ETS approach."

Policy Differences

Since 1990, the National Center for Educational Outcomes, based at the University of Minnesota -Twin Cities, has been tracking accommodations policies for students with disabilities. It found that while the read-aloud accommodation on test reading passages may be controversial, it is not rare. Several states currently allow text passages to be read aloud to students with certain disabilities, though in some cases, that accommodation results in the test being invalidated for accountability purposes.

Kentucky is an example of a state with an expansive read-aloud policy, "on the premise that the intent of reading is to measure comprehension." Hawaii's guidelines for the read-aloud accommodation are much more strict: It allows the use of text-to-speech on its online assessments, but only if the student is a nonreader who will never be able to read any words at any grade level throughout the student's

Patchwork of Policies

States currently vary on whether they allow text passages on state tests to be read aloud to students. Some prohibit this accommodation, while others allow it in certain circumstances, such as for students who are blind or visually impaired. Some states allow it with implications for scoring; for example, the test may be invalidated.

lifetime. The student must also receive all printed material for every subject in an audio format, at all times.

Somewhat more widely used were read-aloud accommodations for test instructions and for math tests, the center found. PARCC and Smarter Balanced have both decided to leave that particular use of read-aloud accommodations up to the school-based special education team working with a given student.

Other variables are whether states allow human readers, software-based text-to-speech, prerecorded audio, or some combination of these.

SOURCES: National Center for Educational Outcomes; Education Week

In allowing the use of read-aloud accommodations, even with some caveats or limitations, the common-core-testing consortia are striking a different path from the National Assessment of Educational Progress, which examines a nationally representative sample of students at grades 4, 8, and 12. NAEP allows a wide range of accommodations, but does not allow passages to be read aloud to students. That restriction has led to large student exclusion rates in some states.

The National Center for Educational Outcomes has suggested that the common-core assessments offer an opportunity for states to develop a coherent, and in its view necessary, multistate policy.

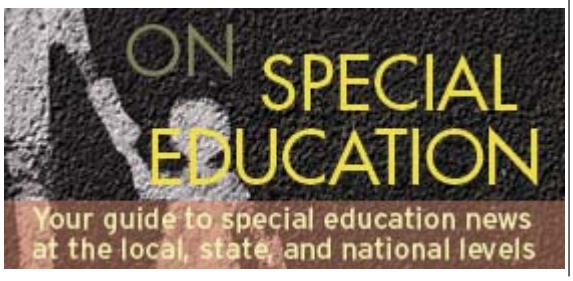
Such a policy should take into account that some students may be able to draw inferences from a text, or answer questions based on its main idea, even if they struggle with translating letters into sounds, also known as decoding. Requiring those students to decode to answer a question based on text comprehension is an "artificial barrier," **according to a report** researchers at the center wrote as guidance for the testing consortia.

"Ensuring that common standards have addressed accessibility concerns does not mean lowering the standards. It does mean, for example, providing a way for students who cannot hear to demonstrate their 'listening' skills; for students who cannot see to demonstrate their 'viewing' skills; and for students who cannot decode to demonstrate their comprehension skills in reading," the report says. But it's not so easy to separate the tasks of reading comprehension and decoding, said literacy expert Timothy Shanahan, the chairman of the department of curriculum and instruction at the University of Illinois at Chicago.

"Part of the task of reading and learning to read is learning to get the words off the page while you think about them. Not having to get the words off the page gives a measurable advantage in most studies," he said.

In a blog post, Mr. Shanahan gave PARCC's read-aloud accommodation decision the "Lindsay Lohan Award for Bad Judgment," a cheeky reference to the troubled starlet. Smarter Balanced made a better move by limiting the read-aloud accommodation in the early grades, he said, but his preference would be not to use it at all.

RELATED BLOG



Visit this blog.

"Part of my concern is I have absolutely no doubt there are going to be school districts and states where people are going to be shopping to get the highest scores they can get," he said. And in states where read-aloud accommodations are used liberally, "those kids are going to do better if they're read to than if they have to read it themselves."

Diane Cordry Golden, the policy coordinator for the Missouri Council of Administrators of Special Education and the project coordinator for the Association of Assistive Technology Act Programs, has a foot in the special education and the assistive-technology worlds. She criticizes views such as Mr. Shanahan's, saying they are akin to "telling a child who can walk with great difficulty that using a wheelchair is not as good. You need to walk, period," she said. "We can't give you technology to work around it; we've got to fix this thing that we see as not normal in you."

It's a legitimate concern that a read-aloud accommodation could be overused, Ms. Golden added, but "this summative test is not an appropriate mechanism to use to address those issues. If you have kids who are really failing readers because of poor instruction, that needs to be driven by something else other than calling out kids with disabilities."

Moving Forward

The consortia are moving ahead with field-testing on a united front, though some state education leaders, such as in Colorado, a PARCC state, registered concerns about allowing the accommodation when the final vote was taken.

"Read-aloud is one of our major concerns. It's pretty fundamental to us," said Robert Hammond, Colorado's commissioner of education, at that meeting, held in June.

But both consortia say it's important to have some policy in place in time for the field-testing. If the accommodation is overused or badly deployed, the field tests will reveal those problems.

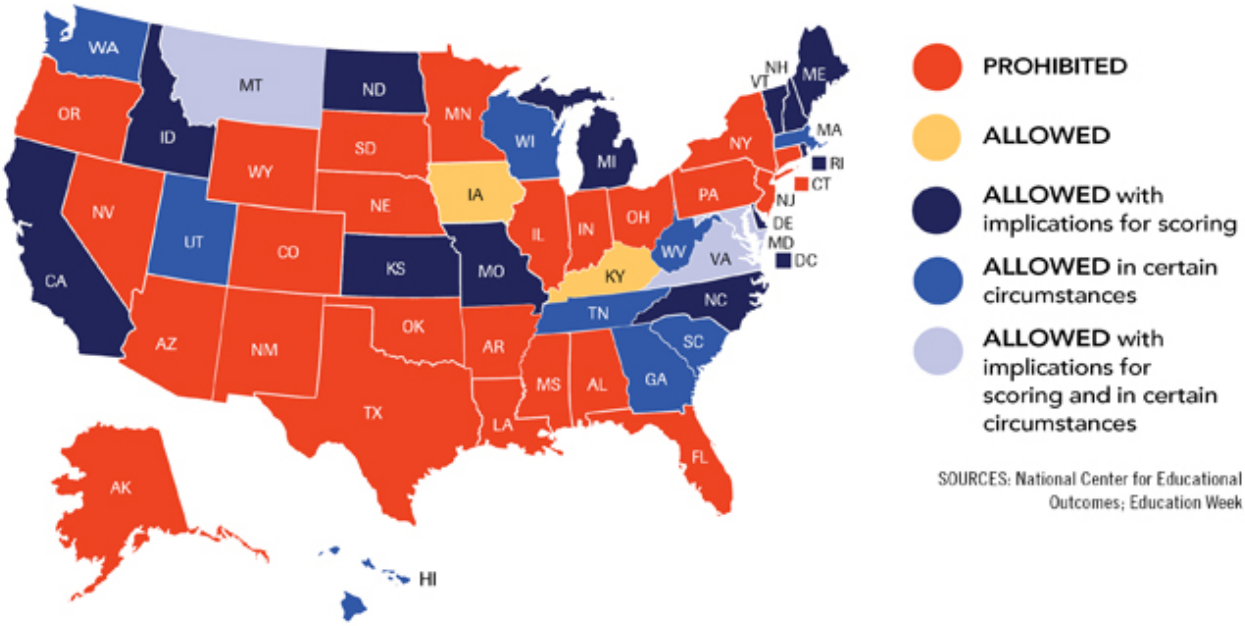
Mitchell D. Chester, the commissioner of education in Massachusetts, a state that does use a read-aloud accommodation on reading tests in limited circumstances, voted for the policy for the purpose of "getting it out into the bloodstream."

"I don't think there's a right or wrong answer to this," said Mr. Chester, the chairman of the PARCC governing board. "I think it's more a philosophical question."

Vol. 33, Issue 10, Pages s20,s21,s22

PATCHWORK OF POLICIES

States currently vary on whether they allow text passages on state tests to be read aloud to students. Some prohibit this accommodation, while others allow it in certain circumstances, such as for students who are blind or visually impaired. Some states allow it with implications for scoring; for example, the test may be invalidated. Other variables are whether states allow human readers, software-based text-to-speech, prerecorded audio, or some combination of these.



SOURCES: National Center for Educational Outcomes; Education Week

FR-13-013

Evaluation of Effects of Implementing A NAEP Mathematics Calculator Booklet Accommodation

Final Report

Prepared for:	**U.S. Department of Education** **National Center for Education Statistics** **Seventh and D Streets, SW** **Washington, DC 20202**
Prepared under:	**MOBIS Contract No. GS-10F-0087J** **E-D08-DO-0081** **Task Order 0036**
Authors:	**Monica A. Gribben** **Tirso E. Diaz**
Date:	**May 28, 2013**

HumRRO
Human Resources Research Organization

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 700, Alexandria, Virginia 22314-1578
Phone: 703.549.3611 | Fax: 703.549.9661 | www.humrro.org

HumRRO
Human Resources Research Organization

Evaluation of Effects of Implementing a NAEP Mathematics Calculator Booklet Accommodation

Table of Contents

List of Tables

Table of Contents (Continued)

List of Figures

Table of Contents (Continued)

Evaluation of Effects of Implementing a NAEP Mathematics Calculator Booklet Accommodation

This study provides a detailed look at the effects of providing a calculator accommodation to students with disabilities (SD) who would otherwise not participate in the National Assessment of Educational Progress (NAEP). The purpose of this study is to determine how increased use of calculator booklets by SD would affect the quality of NAEP score reports. The analyses contained in this report provide information on the effects on reporting of offering calculator accommodation booklets, particularly at the demographic subgroup level. The analyses should be particularly relevant for 2013 reporting because the 2013 NAEP mathematics assessment included calculator accommodation booklets for the first time in NAEP history.

We used 2009 operational NAEP as a baseline and data from the 2011 NAEP inclusion study for the data analyses presented in this report. In 2011, approximately 900 grade 8 students participated in a NAEP special study providing a calculator accommodation (Educational Testing Service [ETS], 2012).[1] Although calculators are used in NAEP, they are only allowed on designated calculator blocks which are assigned randomly to students. Students who would have been excluded from NAEP in 2011 because they used a calculator as an accommodation on their state mathematics assessment and NAEP did not allow a calculator accommodation were offered a calculator booklet to participate in the special study. The calculator booklets were operational booklets with two calculator blocks. In 2011, there were three grade 8 calculator blocks (e.g., A, B, C) as part of the operational assessment. These three blocks were used to create three pairs of blocks for three different calculator booklets (e.g., AB, AC, BC).

This report includes background history and information on the NAEP calculator accommodation issue. Description of the data analyses, results, and recommendations also are presented.

Background

The National Assessment of Educational Progress (NAEP) aims to represent all students. However, students with disabilities (SD) sometimes are excluded from participating because NAEP does not offer an accommodation students use on state assessments (e.g., read-aloud on reading, calculators on mathematics). Inclusion rates vary by jurisdiction, grade, subject area, and subgroup. In 2011, inclusion on NAEP was high with the overall inclusion rate for grade 8 in mathematics at 97 percent. The inclusion rate for SD was lower at 80 percent for grade 8 in mathematics (National Center for Education Statistics [NCES], 2011). The higher the participation rate of all students, the more accurate score reports will be. Bias introduced to NAEP statistics by exclusions, particularly differential exclusion rates, can affect conclusions related to the statistical significance of changes over time and differences among jurisdictions and subgroups. Put another way, excluding students from NAEP assessments because their accommodations are not allowed on NAEP is likely to increase overall scores because scores for SD are generally lower than scores for students without disabilities. For example, in 2011, the mean scale score for SD in grade 8 in mathematics was 250 compared to 288 for students without disabilities (NCES, 2011). Further, SD exclusion effects are becoming increasingly important as NAEP sample sizes increase and greater precision (i.e., smaller standard errors) is achieved by the program (Haertel, 2003). Despite sustained efforts by the U.S. Department of Education to increase the number of SD included in NAEP, high percentages of SD continue to

[1] Approximately 600 grade 4 students also participated in the inclusion study with calculator booklets.

be excluded in some jurisdictions, particularly Maryland, Oklahoma, and Texas where less than 60 percent of SD were included in grade 8 NAEP mathematics in 2011 (NCES, 2011).

NAEP Inclusion Policy

In March 2010, the National Assessment Governing Board (NAGB) revised the NAEP inclusion policy to improve the comparability of score reports of SD across states, districts, and other subgroups. One source of the variation in exclusion rates has been differences in rates of students defined as SD and differences in accommodation allowances across jurisdictions. To reduce variation in exclusion related to these causes, the policy statement (NAGB, 2010) included the following stipulations:

1. Only students with Individualized Education Programs (IEPs) or Section 504 plans can be considered for an accommodation.
2. If the student's IEP or Section 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.
3. If a student's IEP or Section 504 plan does not address a need for an accommodation, school staff are encouraged to include the student without accommodations.
4. If the student requires an accommodation or modification that is not allowed by NAEP, the student is encouraged to take NAEP without it.
5. If a student takes an alternate assessment with *modified* achievement standards, that student should participate on NAEP with the NAEP allowable accommodations on his/her IEP.
6. If a student takes an alternate assessment with *alternate* achievement standards, that student is eligible for exclusion from NAEP. These SD are the only students eligible for exclusion.

The revised policy statement added guidance to the previous policy to encourage school staff to allow students with accommodations or modifications that are not permitted by NAEP to take NAEP without those accommodations (points 4 and 5). Further, the new policy statement explicitly noted that the only students eligible for exclusion are those who take an alternate assessment with alternate achievement standards (part of point 6). Other points of the revised policy were part of the prior policy statement on inclusion.

While most of the requirements are objective, participation by SD who require an accommodation not allowed by NAEP still depends on state and local policies, procedures, and interpretation. In states that allow a particular accommodation on state tests that is not allowed on NAEP, NAEP assessment staff have discussions with school personnel prior to the assessment. School personnel make the determination about participation on a student-by-student basis. Comparability of state results are threatened by differences in (a) state use of accommodations not allowed by NAEP and (b) participation decisions by state and local authorities regarding students who use accommodations not allowed by NAEP.

A new consideration for jurisdiction and school personnel when making their decisions is NAGB's new inclusion goals: Ninety-five percent of all selected students will participate in NAEP (with or without accommodations), and 85 percent of SD and ELL students selected for assessment will participate in NAEP (with or without accommodations). States and jurisdictions that don't meet the goals will be noted in results.

Inclusion rates are calculated as:

$$\frac{\text{assessed+absent+refused}}{\text{assessed+absent+refused+excluded}}.$$

National, state, and district results with fewer than 85 percent inclusion of SD and/or English Language Learner (ELL) students participating in NAEP will be noted in upcoming reports releasing results (NAGB, 2010).

NAEP Inclusion Strategies

During the 2011 administration, the National Center for Education Statistics (NCES) implemented strategies to increase the inclusion rate. In one special study, assessment staff encouraged school personnel to include students who require access to a calculator when responding to mathematics items. Since this accommodation was not allowed on NAEP in 2011 but often allowed on state and other testing, it is a main contributor to differential participation across states and districts. Seventeen states at grade 8 that allow calculator accommodations on state assessments and in other testing situations took part in the study.[2] Other states either did not offer calculator accommodations on their state assessments and/or did not permit calculator use on state assessments. Sampled students who had been excluded from NAEP for the lack of the calculator accommodation were provided with "calculator booklets" (including two calculator blocks). Their scores were not included in calculations for official score reports.[3] The calculator booklet special study was conducted to gather data about how many more students would be included using the strategy and about the performance of this group of students who historically have not participated in NAEP mathematics assessments. NCES wanted to ensure that students requiring calculators on mathematics assessments could meaningfully participate in NAEP.

NAEP Calculator Use

In the regular distribution of items, or spiral, approximately one-third of NAEP's grade 8 mathematics items allow students to use a calculator; the other two-thirds of items assess mathematics knowledge and skills without using a calculator (NAGB, 2007). Two blocks of cognitive items are included in each assessment booklet. Students can receive a booklet with (a) a block of items that allows a calculator and one that does not, (b) two blocks that do not allow a calculator, or (c) two blocks that allow a calculator. Calculator blocks and non-calculator blocks are considered of equal difficulty (Accommodations and Inclusion Coordinating Council [AICC], 2012). In 2011, there was a 52 percent chance that a student sampled in the eighth grade would receive a booklet with a calculator block (AICC, 2012).

Within the calculator blocks, items are categorized by the level of benefit a calculator is to the solution. Calculator active items do benefit from a calculator, calculator neutral items can be solved equally well with or without a calculator, and calculator inactive items are not benefited

[2] The 17 states participating in the special study at grade 8 included: Alaska, California, Indiana, Iowa, Louisiana, Maryland, Massachusetts, Montana, Missouri, New Hampshire, New Jersey, North Dakota, Oklahoma, South Carolina, Texas, Utah, and Vermont.
[3] Some students who use calculator accommodations participate in NAEP without the accommodation. They are randomly assigned a booklet allowing calculator use on zero, one, or two blocks. They may use other accommodations as appropriate. These students were not part of the special study.

by a calculator (NAGB, 2007). Based on the mathematics framework (NAGB, 2010), most objectives permit use of a calculator; only some objectives related to computation limit the use of calculators. However, calculators are only permitted on items in calculator blocks.

Calculator Booklet Special Study

To align with the NAEP mathematics framework (NAGB, 2010), the students in the 2011 calculator booklet special study (892 in grade 8) were tested only on "calculator booklets" that included two calculator blocks (Educational Testing Service [ETS], 2012). The blocks/booklets in the calculator study were the same as those included in the operational assessment. The only difference is that students in the special study were assigned a booklet with two calculator blocks, similar to assigning an accommodation booklet (e.g., read-aloud, bilingual). Characteristics of SD who were assessed in the operational assessment and SD who were assessed in the calculator booklet special study were largely similar in the type of state assessment taken (e.g., with accommodations, modified, alternate) and degree of student disability (ETS, 2012).

NAEP's Design, Analysis, and Reporting (DAR) contractor, ETS, conducted analyses using 2011 NAEP mathematics data that included students who took part in the calculator booklet special study. ETS found that an additional 0.4 percent of grade 8 students could be assessed using a calculator "accommodation" booklet when considering the entire national reporting sample (ETS, 2012). At the national level, exclusion rates would have decreased from 2.5 percent to 2.0 percent for grade 8. In several states use of the calculator booklet would have had a particularly large impact on exclusion rates. Three states at grade 8 would have made the goal of 95 percent inclusion only if the calculator accommodation was part of the operational assessment (Maryland, Oklahoma, and Texas). In Maryland, the exclusion rate would have dropped from 6.3 percent to 2.7 percent for grade 8 students. In Oklahoma, the exclusion rate would have declined from 9.8 percent to 3.3 percent for grade 8 students. In Texas, the exclusion rate would have decreased from 5.2 percent to 3.5 percent for grade 8 students. In other jurisdictions, the option of a calculator accommodation had little, if any, noticeable impact on exclusion rates (e.g., Missouri, New Hampshire, Vermont, and California). The option for a calculator accommodation had an even larger impact on exclusion rates on some Trial Urban Districts Assessment (TUDA) districts (e.g., Baltimore).

Predictably, the effects of a calculator accommodation option on inclusion are amplified when examining the SD subgroup (as compared to all students). The percentage of SD excluded would have decreased from 19.6 percent to 16.0 percent for grade 8 nationwide. At the state level, the percentage of SD excluded would have decreased from 60.1 percent to 19.3 percent for grade 8 in Oklahoma (NCES, embargoed data).

ETS-DAR examined the impact on performance by comparing differences in scale scores with and without students who participated in the calculator booklet special study. Across the national subgroups, scale scores were about 0.3 points greater at grade 8 when the students who participated in the special study were not included in the sample (ETS, 2012). For the SD subgroup, scale scores were about 0.6 points greater at grade 8 when those who participated in the calculator booklet special study were not included in the sample. For grade 8, scale scores without the special study students were about 2.1 points greater in Maryland, 3.5 points greater in Oklahoma, and 6.0 points greater in Baltimore (ETS, 2012). These data provide information on the relative magnitude of the differences in subgroup scale scores.[4]

[4] Statistical significance of the scale score differences was not reported (ETS, 2012).

NAEP Calculator Accommodation

NCES, ETS, members of the November 2011 NAEP Quality Assurance Technical Panel (QATP), and others had concerns that aspects of the NAEP Mathematics Framework may not be equivalent in the NAEP calculator block item pool as compared to the entire mathematics operational item pool. There was concern for the validity of inferences based on scores where a large proportion of students in certain jurisdictions or subgroups do not complete parts of the framework. For example, items that evaluate computational skills are not included in the calculator booklets. In May 2012, ETS-DAR examined framework coverage of the calculator blocks as compared to all blocks. At the general content specification level, ETS researchers found that the calculator blocks were fairly representative and proportional to the framework of the corresponding grade. However, at the more specific objective level, the analysis showed that only 39 of 101 (39 percent) of grade 8 objectives were assessed in the calculator blocks (Freund, 2012). When considering all blocks, almost 80 percent of grade 8 objectives were assessed (80 of 101) (Freund, 2012).

Following a May 2012 meeting of the AICC, the general consensus of staff at NCES and the NAEP Contractor's Alliance was that a booklet with two calculator blocks was not a viable accommodation option for the 2013 assessment (AICC, 2012). While NAGB's goal of higher participation rates and other benefits (e.g., school testing coordinator satisfaction, less social stigma for excluded students) were achieved using the calculator accommodation strategy, lack of content coverage was a concern. In addition, there was the possibility that if states were allowed to request a calculator accommodation for students whose IEP allows them to use calculators on state assessments, then it is likely many more students would use a calculator accommodation than in the 2011 study. In some states (e.g., Virginia) students with state calculator accommodations were encouraged to participate in NAEP without the calculator as an accommodation. However, if NAEP allowed a calculator accommodation some students would be permitted to use the accommodation who previously would have participated without it. Another concern was the differential impact of a calculator accommodation on state results based on the range of SD exclusion rates at the state level.

Based on the results of the calculator booklet special study and increases in inclusion of SD, NCES approved the use of a calculator accommodation – assignment of an accommodation booklet with two calculator blocks – beginning in 2013. To make the calculator accommodation a more viable option, additional booklets were designated as calculator accommodation booklets. In the 2011 special study, there were only three booklets available at grade 8. The Alliance contractors developed additional calculator booklets to increase framework coverage among students requiring a calculator accommodation. To create additional accommodation booklets allowing use of a calculator, some current non-calculator operational blocks were selected for inclusion in the calculator accommodation booklets. SD assigned the calculator accommodation booklets were allowed to use a calculator on the blocks that are assessed operationally without calculators. Any items in the "new" calculator accommodation booklets assessing objectives that prohibit the use of a calculator will be dropped from analysis. In 2017 and beyond, changes to the Mathematics Framework and administration mode (i.e., computer-based assessment) may affect calculator use.

Research Questions

A number of questions could be asked and examined related to allowing a calculator accommodation on NAEP. Given the decision to offer a calculator accommodation in 2013 and available data from the 2011 inclusion study, we focus on questions that can be answered with currently available data and will inform reporting of future results where some students are permitted a calculator accommodation. We addressed the following research questions.

1. How will administration of calculator accommodation booklets to previously excluded students with disabilities affect student **performance** on NAEP mathematics items?
2. How does use of calculator accommodation booklets affect **reporting** of state and subgroup results, using 2009 as a baseline comparison to 2011 results?
3. How do **profiles** of students using a calculator accommodation compare with profiles of other SD and excluded students?
4. Are there differences in **completion rates** for calculator and non-calculator booklets?
5. Are there differences in student performance by calculator block **item types** (calculator active, inactive, and neutral) for students requiring a calculator accommodation?

Method

This study included three phases:

- Gather and Review Data and Prior Results
- Conduct Analyses
- Report Results and Recommendations

Gather and Review Data and Prior Results

HumRRO obtained NAEP data for the 2009 and 2011 operational assessments. The 2011 data include special variables that identify students excluded from the reporting sample but who participated in the calculator accommodation and inclusion studies, along with their plausible value scores. Data for public school students, where the majority of SD attend, were used. In addition, the authors gathered information about the calculator booklet study, including memoranda from ETS documenting the study and results. In addition, HumRRO received a file from ETS containing information tagging items in 2011 calculator blocks by type of calculator use – active, inactive, or neutral. It is expected that a calculator will be helpful in responding to calculator active items. Calculator inactive items do not benefit from the use of a calculator; these items do not require computation. Calculator neutral items require basic calculation, but it is not expected that a calculator will be helpful to most students in responding to these items.

Student Subgroups

To conduct analyses, NAEP participants were sorted into different groups:

- SD who would have been excluded because NAEP did not offer a calculator accommodation and subsequently participated in the calculator study (SD CalcStudy). These students received two calculator blocks.

- SD who use a calculator accommodation on their state assessment but participated in NAEP without a calculator accommodation (SD Calc). These students were randomly assigned either zero, one, or two calculator blocks.
- SD who do not use a calculator accommodation on their state assessment and participated in NAEP (SD NoCalc). These students were randomly assigned either zero, one, or two calculator blocks.
- Students without disabilities who participated in NAEP (NonSD). These students were randomly assigned either zero, one, or two calculator blocks.

Students in the calculator study (SD CalcStudy) were neither part of operational NAEP nor the reporting sample, but had NAEP scores available from the calculator booklets they used. Excluded students were sampled for NAEP but did not participate in the operational assessment or the special study. There are some demographic, disability, and accommodation data on excluded students, but no NAEP scores. Table 1 provides a list of all student groups by disability status and number of calculator blocks with sample sizes. Figure 1 illustrates the different groups of interest sampled to participate in NAEP.

Table 1. Number of Students by Disability Status and Calculator Block Assignment

Student Subgroup	*Uwt n*	*Wt n*
Reporting Sample		
Students without disabilities		
No calculator block	67,980	1,413,084
1 calculator block	70,647	1,471,271
2 calculator blocks	8,889	183,036
Students with disabilities		
Use calc. accomm. on state assessment		
No calculator block	425	9,009
1 calculator block	339	6,864
2 calculator blocks	141	3,155
Do not use calc. accomm. on state assessment		
No calculator block	8,090	163,730
1 calculator block	6,340	132,334
2 calculator blocks	1,552	32,961
Excluded		
Special Study		
Use calc. accomm. on state assessment		
2 calculator blocks	764	13,430
Excluded SD	3,739	67,263

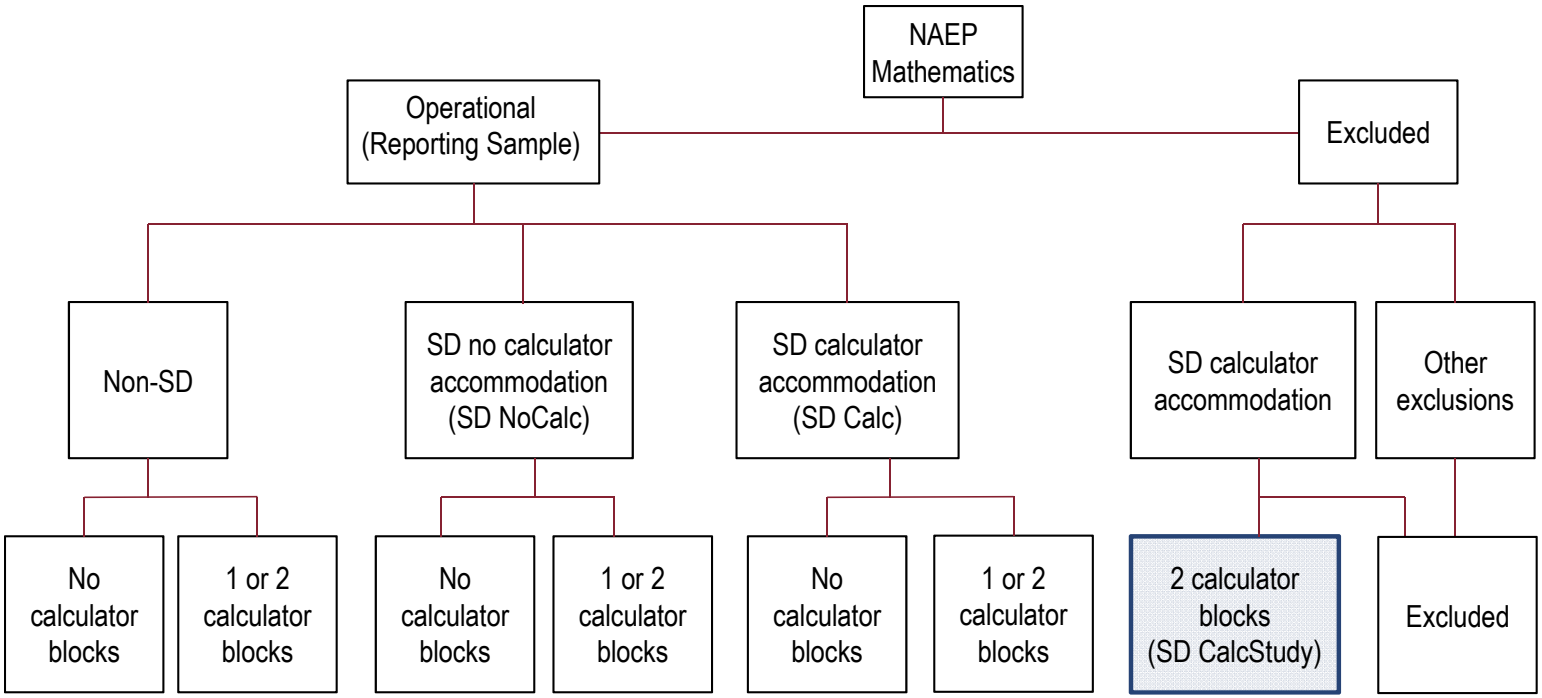


Figure 1. Student Groups in 2011 NAEP Mathematics Sample

Conduct Analyses

We conducted analyses to provide empirical evidence related to each of the study questions using the existing 2009 and 2011 data. In addition, we looked at differences in performance on items by expected calculator activity. We conducted the following analyses:

1. Comparisons of student performance on 2011 NAEP mathematics items for each group of students, including CalcStudy, SD Calc, SD NoCalc, and NonSD.
2. Comparisons of overall and subgroup results from 2009 and 2011 for selected states to examine the effect of including previously excluded students by offering a calculator accommodation.
3. Comparisons of different groups of students on demographic, disability, and accommodation variables based on their participation in NAEP in 2011, including CalcStudy, SD Calc, SD NoCalc, Excluded, and NonSD.
4. Comparisons of the percentage of omitted items and number of items not reached for different groups of students on 2011 NAEP mathematics blocks.
5. Comparisons of percentage correct by expected calculator activity (calculator active, calculator inactive, and calculator neutral) for different types of students.

Results

Effect of Calculator Accommodation on Student Performance

To examine the effect on student performance of offering calculator accommodation booklets to a special subpopulation, we compared mean scale scores (Table 1) and percent correct (Tables 2 and 3) for SD CalcStudy, SD Calc, SD NoCalc, and NonSD.

As shown in Table 2, SD CalcStudy participants achieved a mean scale score of 237.8. SD Calc were assigned booklets containing no calculator blocks, one calculator block (either the first or second block), or two calculator blocks. Their mean scale scores ranged from 235.7 for those who received a calculator block first and a non-calculator block second to 240.9 for those who received two calculator blocks. Some of these subgroups are small and the standard errors in the mean scale scores are large, so we examined the range of mean scale scores for the lower and upper confidence intervals to make comparisons between groups. There is considerable overlap in the confidence intervals of SD CalcStudy and SD Calc. SD NoCalc scored 246.0, slightly higher than SD Calc. There was less overlap between confidence intervals for the SD CalcStudy and SD NoCalc. Students with no disabilities received much higher scores (286.1) than SD, echoing the achievement gap between SD and students without disabilities (Frieden, 2004).

A similar pattern emerged when examining percent correct for the same subgroups of students (see Table 3). SD CalcStudy participants received a *p*-value of 27.0, compared to a range of 24.5 to 28.4 for SD Calc. Figure 1 illustrates the considerable overlap of confidence intervals for all groups of SD requiring calculator accommodations. There was no overlap of confidence intervals for SD NoCalc (29.5-32.1) and those who participated in the calculator study (SD CalcStudy; 25.1-29.0). As seen in the mean scale score comparisons, students without disabilities (NonSD) achieved significantly greater *p*-values than those with disabilities. Results at the block level showed similar patterns (see Table 4).

Table 2. Mean Scale Scores for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

Student Subgroup	Calculator Use		Uwt *n*	Mean	S.E.	95% C.I.
	Block 1	Block 2				
Calculator study (SD CalcStudy)	Y	Y	764	237.8	2.6	[232.7 , 242.8]
SD with calculator accommodation--operational NAEP (SD Calc)	N	N	425	239.2	2.8	[233.8 , 244.6]
	N	Y	138	235.8	4.1	[227.9 , 243.8]
	Y	N	201	235.7	3.6	[228.7 , 242.7]
	Y	Y	141	240.9	5.0	[231.2 , 250.7]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	1,552	246.0	2.4	[241.3 , 250.6]
Non-SD--operational NAEP (NonSD)	Y	Y	8,889	286.1	0.7	[284.8 , 287.4]

Table 3. Overall Percent Correct for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

Student Subgroup	Calculator Use		Uwt *n*	Percent Correct	S.E.	95% C.I.
	Block 1	Block 2				
Calculator study (SD CalcStudy)	Y	Y	764	27.0	1.0	[25.1 , 29.0]
SD with calculator accommodation--operational NAEP (SD Calc)	N	N	425	26.3	1.0	[24.3 , 28.3]
	N	Y	138	24.5	1.3	[22.0 , 27.0]
	Y	N	201	24.6	1.1	[22.4 , 26.8]
	Y	Y	141	28.4	1.5	[25.4 , 31.4]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	1,552	30.8	0.7	[29.5 , 32.1]
Non-SD--operational NAEP (NonSD)	Y	Y	8,889	48.4	0.4	[47.7 , 49.1]

Table 4. Percent Correct by Block for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

Student Subgroup	Calculator Use		Block 1			Block 2		
	Block 1	Block 2	Percent Correct	S.E.	95% C.I.	Percent Correct	S.E.	95% C.I.
Calculator study (SD CalcStudy)	Y	Y	28.6	1.2	[26.2 , 31.0]	25.3	1.0	[23.4 , 27.2]
SD with calculator accommodation--operational NAEP (SD Calc)	N	N	27.9	1.0	[25.9 , 29.9]	24.8	1.2	[22.4 , 27.1]
	N	Y	25.1	1.5	[22.3 , 28.0]	24.3	1.6	[21.1 , 27.5]
	Y	N	23.2	1.2	[20.8 , 25.6]	25.9	1.5	[23.0 , 28.9]
	Y	Y	27.9	1.7	[24.6 , 31.1]	29.0	2.2	[24.7 , 33.3]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	30.2	0.6	[29.1 , 31.4]	31.5	0.8	[29.9 , 33.1]
Non-SD--operational NAEP (NonSD)	Y	Y	48.2	0.4	[47.5 , 49.0]	48.6	0.4	[47.9 , 49.3]

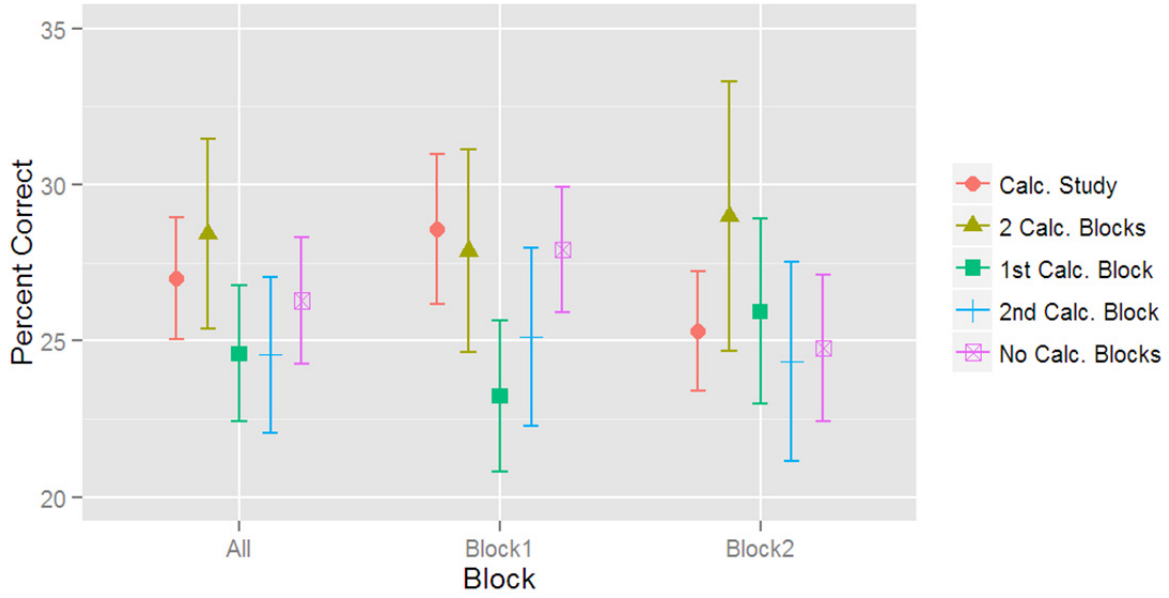


Figure 2. Percent Correct on 2011 NAEP Mathematics (Grade 8) by Block for SD Requiring a Calculator Accommodation.

Effect of Calculator Accommodation on State Reporting

Exclusion Rates

To examine the effect on reporting of including previously excluded students by offering a calculator accommodation on NAEP mathematics, we compared overall and subgroup results from 2009 and 2011. These analyses were conducted for the four states with the greatest number of students participating in the calculator special study (listed in order by initial exclusion rate, starting with the largest): Oklahoma, Maryland, Texas, and New Jersey. By providing a calculator accommodation on NAEP, these states would have been able to reduce their exclusion rates on grade 8 mathematics in 2011:

- Oklahoma – from 9.8 percent to 3.3 percent
- Maryland – from 6.3 percent to 2.7 percent
- Texas – from 5.2 percent to 3.5 percent
- New Jersey – from 4.2 percent to 1.4 percent

State and Subgroup Performance

The state-level results in Tables 5 through 8 show the potential for differences in state and state-level subgroup reporting when a large number of previously excluded SD are included in NAEP reports. In Oklahoma, as shown by the green shading in Table 5a, comparisons of 2009 and 2011 grade 8 results indicate significant gains for all students as well as females and Whites. When including students with calculator accommodations (SD CalcStudy) in 2011, the differences between 2009 and 2011 for all students, females, and Whites are not statistically significant (as shown by the yellow shading). Other subgroups, including males, Blacks, and Hispanics, show non-significant gains – essentially unchanged – when comparing 2009 to 2011 operational results. When SD CalcStudy participants are included, each of these subgroups (males, Blacks, and Hispanics) shows decreases in performance that are not statistically significant.

Including the calculator accommodation students in the 2011 results increased the number of students with disabilities, therefore we adjusted the 2009 results to include a similar proportion of SD who would have participated in NAEP if a calculator accommodation had been available in 2009.[5] These adjustments are shown in Tables 5b through 8b. In Oklahoma, when comparing 2011 data with calculator accommodations students included to 2009 adjusted data, females showed significant gains, but there were no significant changes overall or in other subgroups.

In Maryland, as shown in Table 6a, there were no significant differences between 2009 and 2011 grade 8 results with or without students with calculator accommodations. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 6b), there were still no significant changes overall or in any subgroup.

[5] The 2009 mean scores were adjusted by assuming that the percentage of excluded SD that would have participated if a calculator accommodation was available in 2009 is equal to the percentage of excluded SD who participated in the calculator study in 2011. This was carried out by increasing the weights of SD included in the reporting sample in 2009 so that their total would reflect the number of additional SD from those previously excluded who would have participated with a calculator accommodation. The adjustment assumes that performance of included and excluded SD are comparable.

Table 5a. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009			NAEP 2011 No Calculator Accommodation						NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	40,910	275.7	1.0	36,936	279.2	1.1	3.5	2.3	0.02	39,623	276.0	1.1	0.3	0.2	0.86
Male	20,771	277.6	1.4	17,957	280.0	1.3	2.4	1.2	0.21	19,729	276.0	1.4	-1.6	-0.8	0.43
Female	20,139	273.8	1.2	18,979	278.4	1.2	4.6	2.7	0.01	19,894	275.9	1.2	2.2	1.3	0.20
White	23,875	282.1	1.2	20,292	286.4	1.2	4.3	2.5	0.01	21,668	283.5	1.2	1.4	0.8	0.40
Black	3,890	260.6	2.9	3,879	262.9	2.3	2.2	0.6	0.54	4,121	259.7	2.4	-1.0	-0.3	0.80
Hispanic	4,415	262.6	2.6	4,088	263.6	2.9	0.9	0.2	0.81	4,241	262.5	2.8	-0.2	0.0	0.96

Table 5b. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009 (Adjusted)			NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	42,725	274.2	1.1	39,623	276.0	1.1	1.8	1.2	0.25
Male	21,909	275.9	1.5	19,729	276.0	1.4	0.1	0.0	0.97
Female	20,817	272.4	1.2	19,894	275.9	1.2	3.6	2.1	0.04
White	24,882	280.6	1.3	21,668	283.5	1.2	2.9	1.7	0.10
Black	4,107	259.0	3.0	4,121	259.7	2.4	0.7	0.2	0.86
Hispanic	4,567	261.4	2.7	4,241	262.5	2.8	1.0	0.3	0.79

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

*Difference between 2011 and 2009 means (2011 minus 2009).

Table 6a. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009			NAEP 2011 No Calculator Accommodation						NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	53,951	288.3	1.1	57,086	287.9	1.2	-0.4	-0.3	0.80	59,299	286.1	1.2	-2.3	-1.4	0.17
Male	27,043	289.8	1.5	28,470	289.4	1.5	-0.3	-0.2	0.88	29,968	287.1	1.5	-2.7	-1.3	0.20
Female	26,907	286.9	1.3	28,616	286.4	1.4	-0.5	-0.3	0.79	29,331	285.0	1.4	-1.9	-1.0	0.33
White	26,267	302.9	1.4	25,873	303.3	1.3	0.4	0.2	0.84	26,558	302.0	1.3	-0.9	-0.5	0.62
Black	18,847	265.8	1.4	19,409	266.7	1.3	0.9	0.5	0.64	20,522	264.5	1.2	-1.2	-0.7	0.50
Hispanic	5,155	274.9	2.2	6,066	273.4	3.6	-1.4	-0.3	0.74	6,408	271.9	3.3	-2.9	-0.7	0.46

Table 6b. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009 (Adjusted)			NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	56,192	287.4	1.1	59,299	286.1	1.2	-1.3	-0.8	0.43
Male	28,587	288.5	1.5	29,968	287.1	1.5	-1.4	-0.7	0.51
Female	27,605	286.3	1.3	29,331	285.0	1.4	-1.2	-0.6	0.52
White	27,468	302.0	1.4	26,558	302.0	1.3	0.0	0.0	0.99
Black	19,698	264.9	1.4	20,522	264.5	1.2	-0.3	-0.2	0.86
Hispanic	5,324	273.9	2.2	6,408	271.9	3.3	-2.0	-0.5	0.62

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

*Difference between 2011 and 2009 means (2011 minus 2009).

In Texas, as shown in Table 7a, comparisons of 2009 and 2011 grade 8 results indicated significant gains for all students as well as females and Hispanics. When including students with calculator accommodations in 2011, the difference between 2009 and 2011 for all students was no longer statistically significant. However, females and Hispanics still showed significant gains. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 7b), there were significant gains overall and for females and Hispanics.

In New Jersey, as shown in Table 8a, there were no significant changes between 2009 and 2011 grade 8 results for all students as well as all subgroups. All students and subgroups, except males, showed non-significant gains – essentially unchanged – when comparing 2009 to 2011 operational results. There were non-significant declines for males. When students who participated in the calculator accommodation study were included the results were essentially the same. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 8b), there were no significant changes overall or in any subgroup.

Table 7a. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009			NAEP 2011 No Calculator Accommodation						NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	306,243	286.7	1.3	323,200	290.5	0.9	3.8	2.5	0.01	328,946	289.6	0.9	2.9	1.9	0.05
Male	151,897	287.4	1.4	162,059	290.9	1.2	3.6	1.9	0.06	166,075	289.8	1.2	2.4	1.3	0.20
Female	154,345	286.0	1.4	161,141	290.0	0.9	4.0	2.3	0.02	162,870	289.5	0.9	3.5	2.0	0.04
White	112,194	300.8	1.6	102,792	303.7	1.2	2.9	1.5	0.15	104,305	302.9	1.2	2.1	1.0	0.30
Black	41,356	272.3	1.8	40,946	277.0	1.8	4.7	1.8	0.07	42,702	275.3	1.9	3.0	1.2	0.25
Hispanic	139,625	277.2	1.5	163,463	283.4	1.0	6.2	3.5	0.00	165,937	282.8	1.0	5.6	3.1	0.00

Table 7b. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009 (Adjusted)			NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	311,066	286.2	1.3	328,946	289.6	0.9	3.5	2.2	0.03
Male	154,854	286.9	1.5	166,075	289.8	1.2	2.9	1.5	0.12
Female	156,212	285.5	1.5	162,870	289.5	0.9	4.0	2.3	0.02
White	113,829	300.3	1.6	104,305	302.9	1.2	2.6	1.3	0.21
Black	42,277	271.7	1.8	42,702	275.3	1.9	3.6	1.4	0.18
Hispanic	141,777	276.8	1.5	165,937	282.8	1.0	6.0	3.3	0.00

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

*Difference between 2011 and 2009 means (2011 minus 2009).

Table 8a. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009			NAEP 2011 No Calculator Accommodation						NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	97,691	292.7	1.4	91,430	294.3	1.2	1.6	0.9	0.39	94,097	293.2	1.2	0.5	0.3	0.79
Male	49,178	294.8	1.7	46,263	294.6	1.6	-0.2	-0.1	0.93	47,782	293.2	1.6	-1.6	-0.7	0.50
Female	48,513	290.5	1.5	45,167	294.0	1.2	3.5	1.8	0.07	46,315	293.1	1.2	2.6	1.3	0.18
White	57,578	301.6	1.2	50,955	304.4	1.1	2.9	1.8	0.07	52,683	302.9	1.1	1.3	0.8	0.41
Black	15,572	267.2	2.6	14,600	271.7	2.3	4.5	1.3	0.19	15,092	270.9	2.3	3.7	1.1	0.28
Hispanic	16,658	271.8	2.8	18,069	274.7	2.5	2.9	0.8	0.44	18,451	274.1	2.4	2.3	0.6	0.53

Table 8b. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011)

Student Group	NAEP 2009 (Adjusted)			NAEP 2011 With Calculator Accommodation					
	Wt *n*	Mean	S.E.	Wt *n*	Mean	S.E.	Difference*	*t*	*p*-value
ALL	98,853	292.3	1.4	94,097	293.2	1.2	0.9	0.5	0.63
Male	49,900	294.4	1.7	47,782	293.2	1.6	-1.1	-0.5	0.63
Female	48,953	290.1	1.5	46,315	293.1	1.2	3.0	1.5	0.13
White	58,280	301.2	1.2	52,683	302.9	1.1	1.7	1.1	0.29
Black	15,811	266.7	2.6	15,092	270.9	2.3	4.2	1.2	0.23
Hispanic	16,845	271.4	2.8	18,451	274.1	2.4	2.7	0.7	0.47

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

*Difference between 2011 and 2009 means (2011 minus 2009).

Figures 3 through 6 graphically summarize the differences between 2009 and 2011 mean scale scores. For each subgroup, three error bars are shown representing the 95 percent confidence interval of differences in mean scale scores based on three calculations: (a) using reporting sample results for both 2009 and 2011, (b) including students with a calculator accommodation in 2011 without adjusting 2009 results, and (c) including students with a calculator accommodation in 2011 while at the same time adjusting 2009 results. The lower and upper endpoints of the error bars represent the bounds of the confidence interval, with the midpoint representing the mean scale score difference. Error bars entirely above or below the zero horizontal reference line indicate differences between 2009 and 2011 results that were statistically significant.

As shown in Figure 3, the error bars representing operational scale score differences for Oklahoma were above the zero reference line for all students, females, and Whites, indicating statistically significant gains. All three error bars then shifted downward, enclosing the reference line, indicating not statistically significant gains when including students with calculator accommodations. After adjusting the 2009 results, all three error bars shifted upward but only the error bar representing females was above the reference line. For all other groups, the error bars followed a similar pattern of shifting downward when including students with a calculator accommodation, and then shifting upward after adjusting the 2009 results, but all indicated not statistically significant gains.

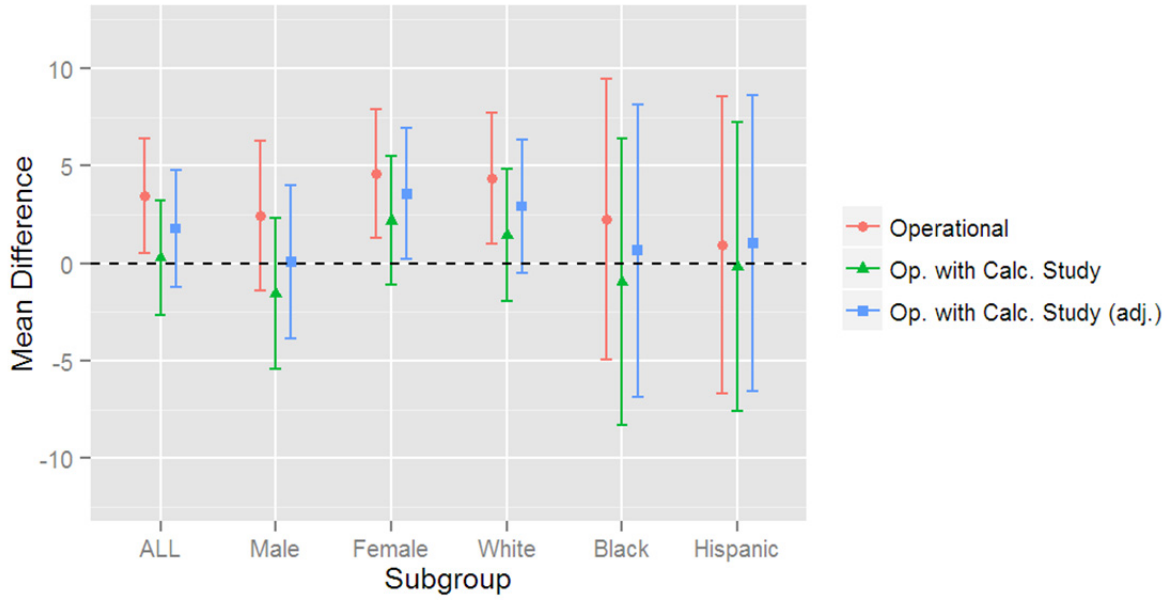


Figure 3. Subgroup Mean Differences (2011-2009) for Oklahoma grade 8 students.

Figure 4 graphically summarizes the differences between 2009 and 2011 mean scale scores for Maryland. Again, all error bars followed the general pattern of shifting downward when including students with calculator accommodations, and then shifting upward after adjusting the 2009 results across all subgroups. However, all error bars indicated not statistically significant loses or gains. Less dramatic changes were observed in Maryland compared to those observed in Figure 3 for Oklahoma.

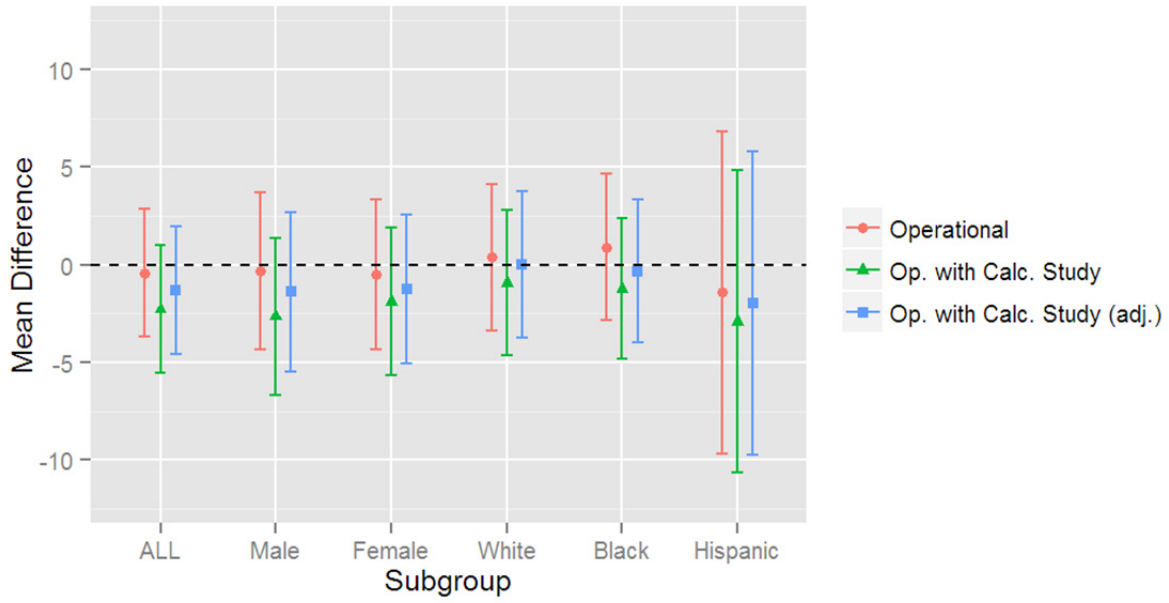


Figure 4. Subgroup Mean Differences (2011-2009) for Maryland grade 8 students.

Figure 5 graphically summarizes the differences between 2009 and 2011 mean scale scores described above for Texas. Changes in the location of error bars relative to the reference line were very minimal across all subgroups. Note that error bars for Hispanic students were narrower compared to corresponding error bars in Figures 2 and 3, as expected, since Texas has a relatively higher proportion of Hispanic students than Oklahoma and Maryland.

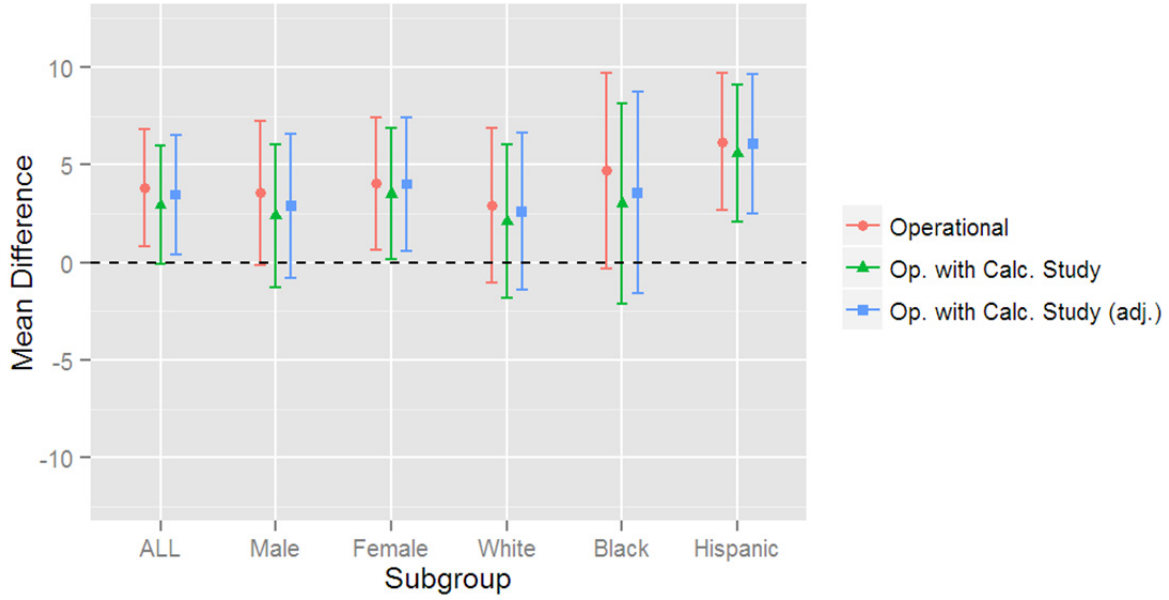


Figure 5. Subgroup Mean Differences (2011-2009) for Texas grade 8 students.

Figure 6 graphically summarizes the differences between 2009 and 2011 mean scale scores for New Jersey. All error bars indicated not statistically significant gains or losses.

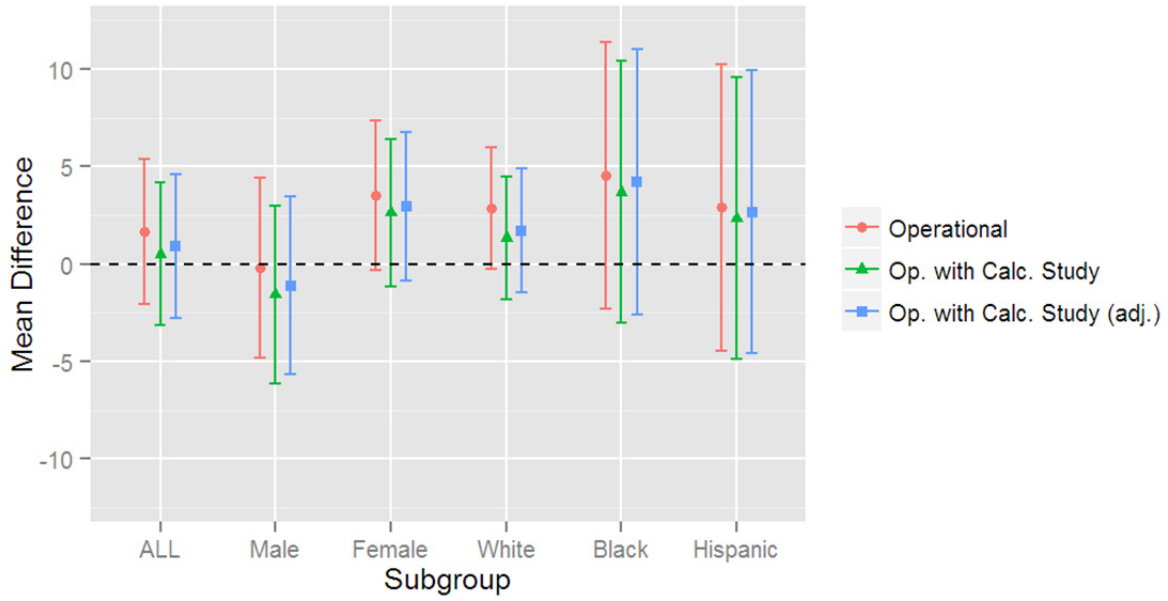


Figure 6. Subgroup Mean Differences (2011-2009) for New Jersey grade 8 students.

Profiles of Students

To put differences in student performance and the effect of a calculator accommodation on reporting in context, we examined characteristics of students to develop descriptive profiles using demographic variables – gender, race, ethnicity, and participation in the National School Lunch Program (NSLP). In addition, distributions of disability and accommodations were examined. To describe students with disabilities, we used five measures:

- Degree of disability – mild, moderate, and profound.
- Grade level of performance – performing at or above grade level, one year below grade level, and two or more years below grade level.
- Type of disability – five most frequent disabilities: specific learning disability, speech impediment, emotional disturbance, autism, and mental retardation.
- Type of accommodation – most frequently used accommodations: extended time, small group, read all in English, read directions in English, read occasionally in English, breaks, cueing, school staff administer test, one on one, special equipment/preferential seating, and scribe.
- Number of accommodations – three or more.

Demographic Characteristics

As shown in Table 9, students participating in the special study were more likely to be male and Black than non-SD. They also were more likely to participate in the NSLP than SD and non-SD included in operational NAEP.

Table 9. Demographic profiles of grade 8 students participating in NAEP mathematics items (2011)

Student Group	Calculator Study (SD CalcStudy)		Included SD (SD Calc and SD NoCalc)					Excluded SD					Included Non-SD (NonSD)				
	Percent	S.E.	Percent	S.E.	Difference*	*t*	*p*-value	Percent	S.E.	Difference*	*t*	*p*-value	Percent	S.E.	Difference*	*t*	*p*-value
ALL	100.0	0.0	100.0	0.0	0.0	.	.	100.0	0.0	0.0	.	.	100.0	0.0	0.0	.	.
Male	65.4	2.7	64.8	0.6	-0.6	-0.2	0.83	64.8	1.1	-0.6	-0.2	0.85	49.0	0.1	-16.3	-6.1	0.00
Female	34.6	2.7	35.2	0.6	0.6	0.2	0.83	35.2	1.1	0.6	0.2	0.85	51.0	0.1	16.3	6.1	0.00
White	43.3	2.8	53.0	0.6	9.7	3.3	0.00	49.8	1.4	6.5	2.1	0.04	53.7	0.3	10.4	3.6	0.00
Black	25.8	2.9	18.8	0.5	-7.0	-2.4	0.02	24.5	1.1	-1.3	-0.4	0.68	15.3	0.3	-10.5	-3.6	0.00
Hispanic	23.3	3.5	22.3	0.6	-1.1	-0.3	0.76	20.7	1.5	-2.7	-0.7	0.48	22.5	0.3	-0.8	-0.2	0.82
NSLP	64.6	3.0	58.0	0.6	-6.6	-2.1	0.03	63.2	1.3	-1.5	-0.4	0.66	46.5	0.3	-18.2	-6.0	0.00

*Difference with SD Calculator Accommodation

Disability Characteristics

Data in Table 10 Indicate most students with disabilities have a mild or moderate disability. Of those participating in operational NAEP in 2011, half had a mild disability compared to 46 percent of students participating in the calculator study. More students in the calculator study were considered working at one year below grade level (30 percent) than SD in NAEP (26 percent). The gap was even greater for students performing two or more years below grade level. Nearly 37 percent of SD receiving calculator accommodations were two or more years below grade level compared to approximately one-third of SD participating in operational NAEP without a calculator accommodation.

Table 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011)

	Calculator Study (SD CalcStudy)		Included SD (SD Calc and SD NoCalc)					Excluded SD				
Student Group	Percent	SE	Percent	S.E.	Difference*	*t*	*p*-value	Percent	S.E.	Difference*	*t*	*p*-value
Degree of Disability												
Profound	5.7	1.2	6.1	0.3	0.4	0.3	0.73	29.8	1.4	24.1	13.0	0.00
Moderate	35.2	3.5	34.7	0.7	-0.5	-0.1	0.90	34.1	1.5	-1.1	-0.3	0.78
Mild	45.7	2.7	50.0	0.8	4.3	1.5	0.13	23.2	1.3	-22.5	-7.4	0.00
Grade Level Student Performs on												
At or above	18.5	3.5	28.6	0.7	10.1	2.9	0.00	9.3	1.0	-9.3	-2.6	0.01
1 year below	29.6	3.3	25.7	0.8	-3.8	-1.1	0.26	16.2	0.9	-13.3	-3.9	0.00
2 or more years below	36.9	3.6	32.7	0.8	-4.2	-1.1	0.25	52.0	1.6	15.1	3.8	0.00

*Difference with SD Calculator Accommodation

Figure 7 shows that SD are more likely to be male than female. As shown in Figure 8, students in the calculator study were slightly more likely to be a minority. Students in the calculator study were slightly more likely to participate in the NSLP than SD in NAEP and non-SD (Figure 9). There was little difference between excluded SD and calculator study students on participation in the NSLP.

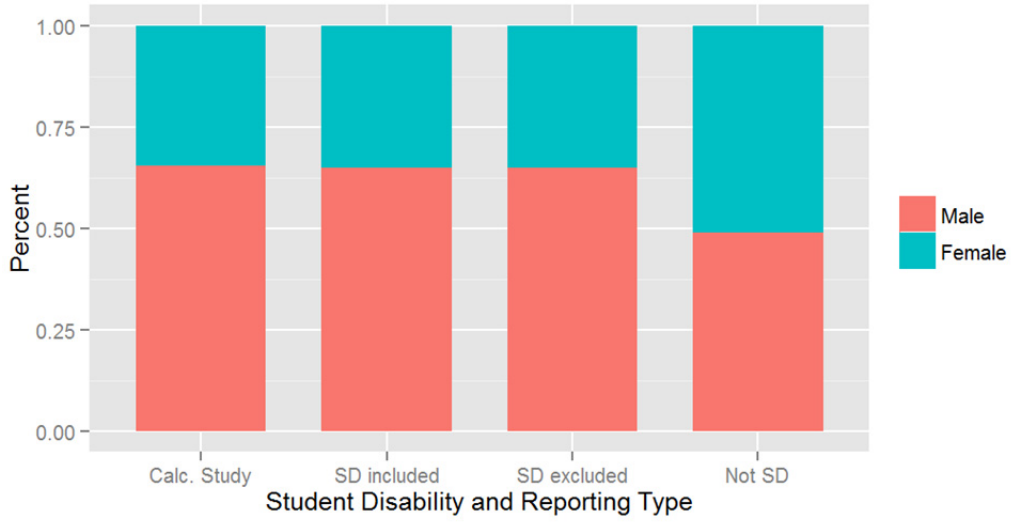


Figure 7. Demographic profiles by gender of grade 8 students participating in NAEP mathematics (2011)

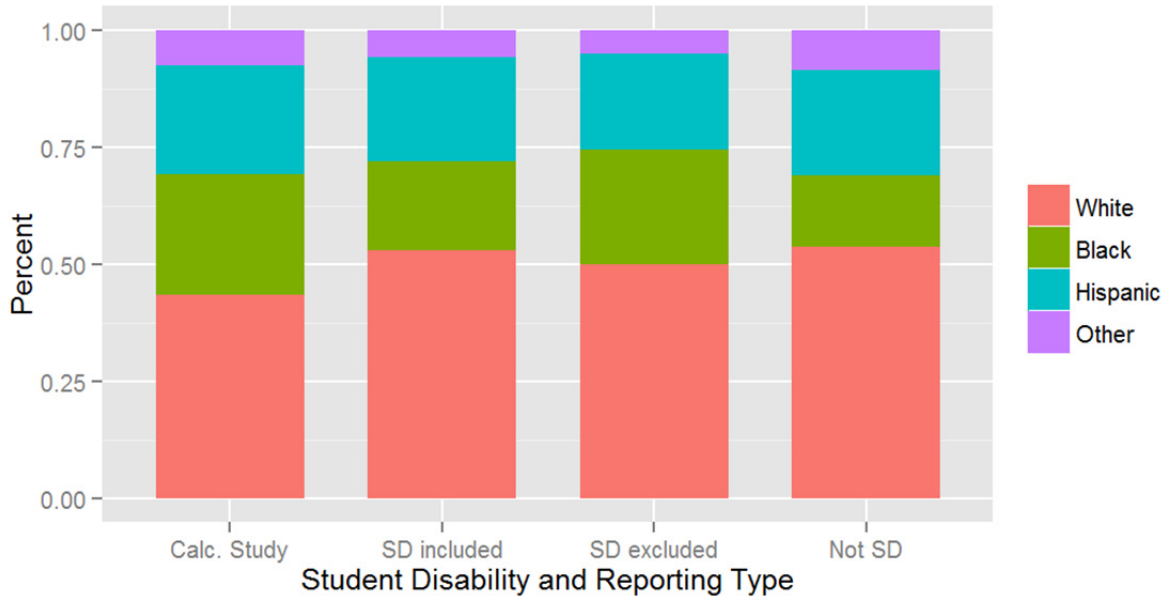


Figure 8. Demographic profiles by race/ethnicity of grade 8 students participating in NAEP mathematics (2011)

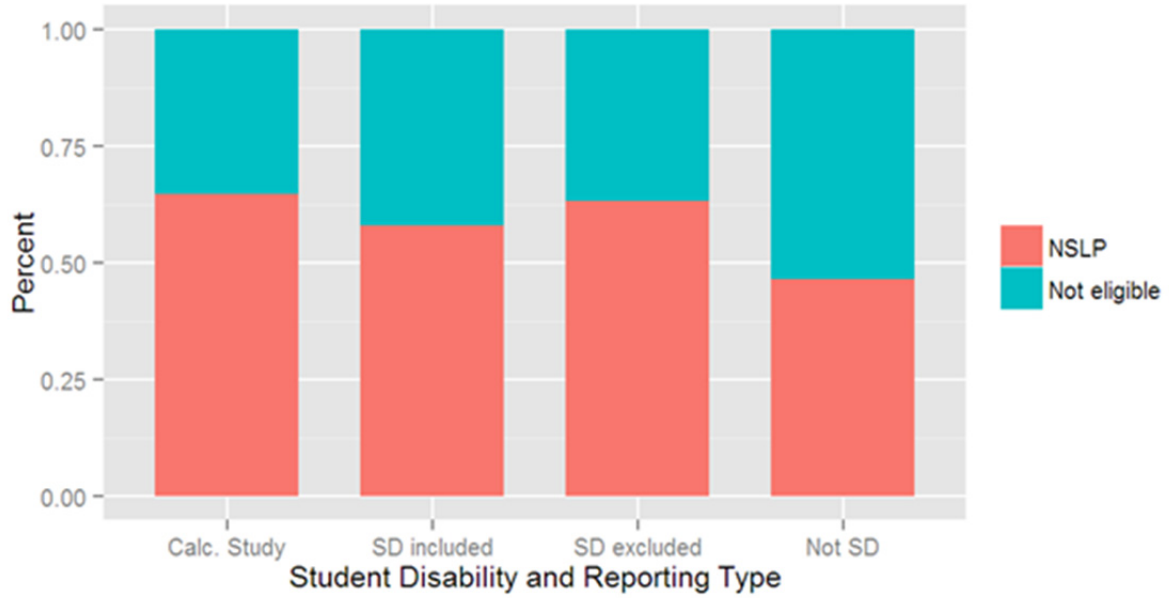


Figure 9. Demographic profiles by participants in the National School Lunch Program of grade 8 students participating in NAEP mathematics (2011)

Figure 10 visually shows students excluded from NAEP were much more likely to have profound disabilities and much less likely to have mild disabilities than SD who participated in NAEP, either operationally or as part of the calculator study. When looking at grade level (Figure 11), calculator study students were less likely to be at or above grade level and slightly more likely to be 2 or more years below grade level than SD in operational NAEP. SD excluded from NAEP were much more likely to be 2 or more years below grade level than other SD.

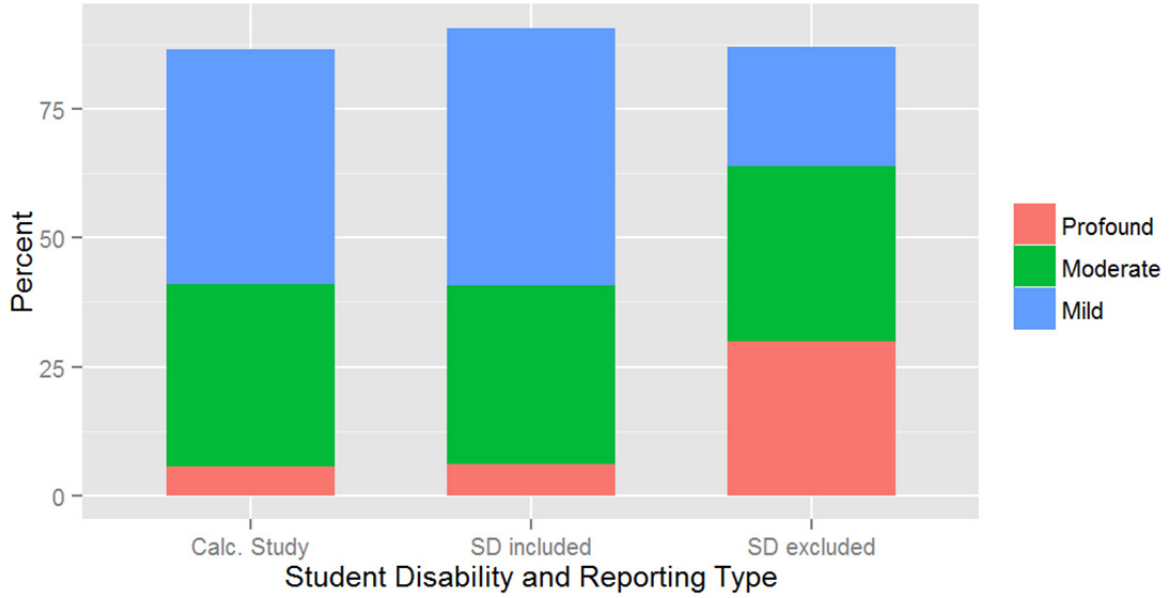


Figure 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011)

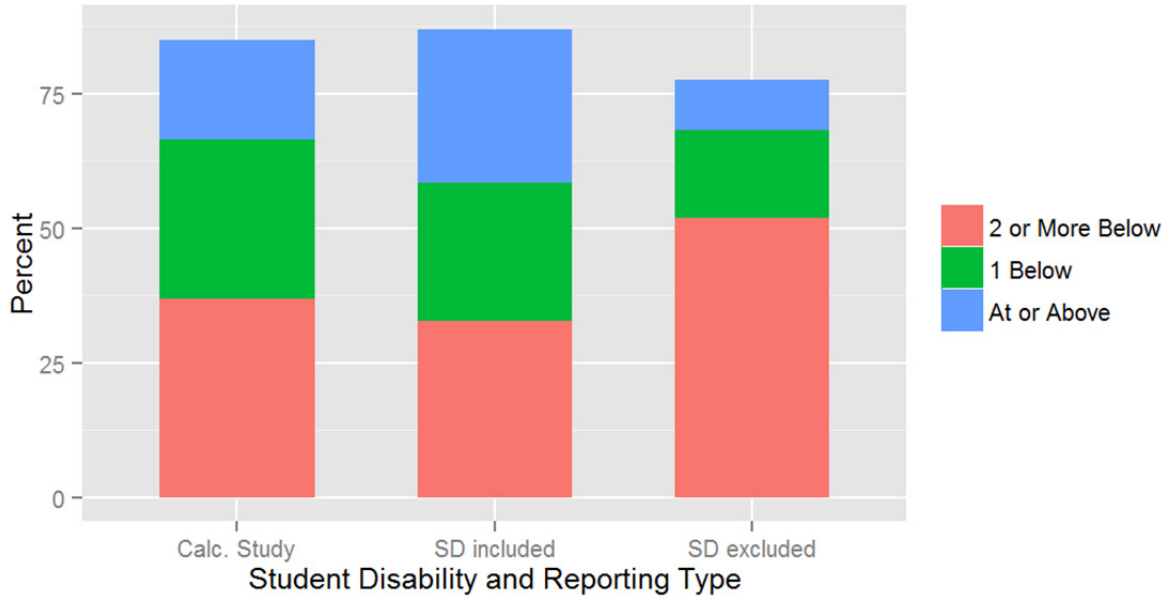


Figure 11. Grade level of performance of grade 8 students participating in NAEP mathematics (2011)

In Table 11 and Figure 12 students participating in the calculator study with a calculator accommodation were similar to SD participating in operational NAEP in terms of the types of disabilities they have. Excluded students were significantly less likely to have a specific learning disability and more likely to have a specific disability such as mental retardation, autism, visual impairment, developmental delay, or brain injury than students in the calculator study.

As shown in Table 12 calculator study students were more likely to need read all or occasionally in English, preferential seating, and a scribe compared to SD participating in NAEP operationally.

Figure 13 illustrates the differences between SD in the calculator study and operational NAEP. There was overlap on many of the accommodations, showing little difference between the two groups except for “read aloud occasionally” and “read aloud all.”

Data in Table 13 show the percentage of students receiving one or more accommodations, two or more accommodations, and so forth. Students with disabilities participating in operational NAEP used significantly fewer accommodations than students in the calculator study. Note, all students in the calculator study received a calculator accommodation and three-quarters received at least two other accommodations, often small group and extended time. More than one in five students in the calculator study received five or more accommodations compared to less than 4 in 100 SD in operational NAEP.

Table 11. Type of disability of grade 8 students participating in NAEP mathematics (2011)

Type of Disability	Calculator Study (SD CalcStudy)		Included SD (SD Calc and SD NoCalc)					Excluded SD				
	Percent	SE	Percent	SE	Difference*	*t*	*p*-value	Percent	SE	Difference*	*t*	*p*-value
Any disability	97.4	1.5	100.0	0.0	2.6	1.7	0.09	100.0	0.0	2.6	1.7	0.09
Specific learning disability	64.1	2.9	58.2	0.7	-5.8	-1.9	0.05	40.3	1.4	-23.7	-7.2	0.00
Other health disability	15.6	2.3	15.1	0.5	-0.5	-0.2	0.83	12.3	0.9	-3.3	-1.3	0.18
Speech impairment	6.7	1.4	9.1	0.4	2.4	1.6	0.10	10.5	0.8	3.8	2.3	0.02
Emotional disturbance	5.1	1.2	5.2	0.3	0.1	0.1	0.94	6.4	0.6	1.3	1.0	0.34
Autism	2.9	1.2	3.5	0.2	0.7	0.5	0.60	8.5	0.7	5.6	4.0	0.00
Mental retardation	2.4	0.8	2.5	0.2	0.1	0.2	0.85	24.1	1.1	21.7	15.9	0.00

*Difference with SD calculator accommodation

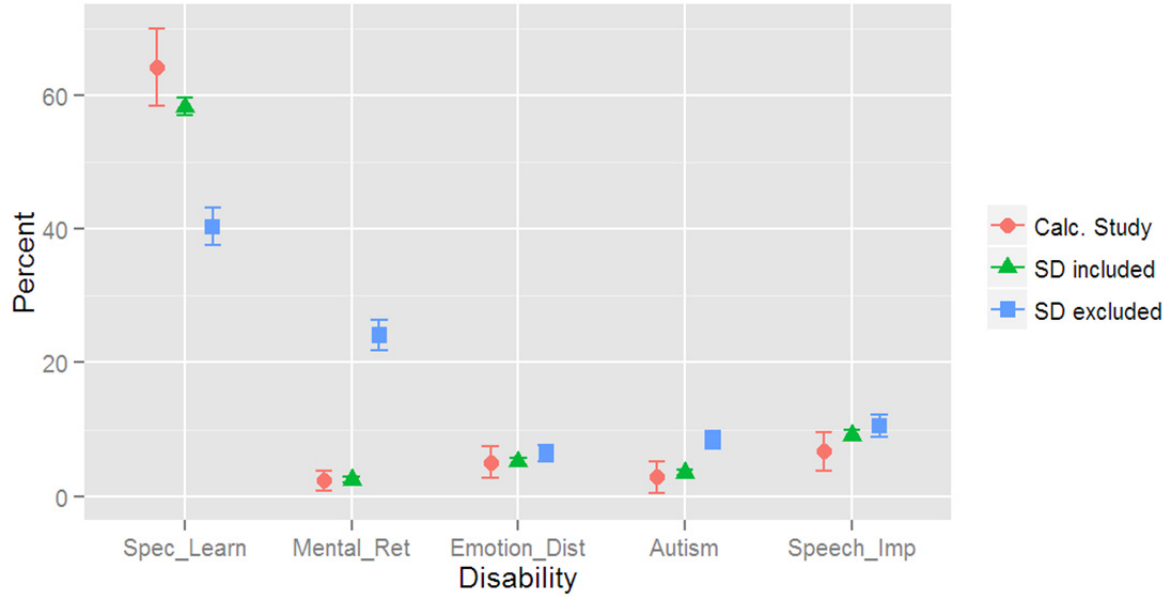


Figure 12. Type of disability of grade 8 students participating in NAEP mathematics (2011)

Table 12. Types of accommodations used by grade 8 students participating in NAEP mathematics (2011)

Type of Accommodation	Calculator Study (SD CalcStudy)		Included SD (SD Calc and SD NoCalc)				
	Percent	SE	Percent	SE	Difference*	*t*	*p*-value
All accommodations	100.0	0.0	83.6	0.5	-16.4	-31.3	0.00
Extended time	59.7	4.7	68.3	0.8	8.6	1.8	0.07
Small group	66.1	3.2	64.5	0.8	-1.6	-0.5	0.62
Read all in English	44.6	4.2	26.5	0.6	-18.1	-4.3	0.00
Read directions in English	15.0	2.3	19.7	0.7	4.6	1.9	0.05
Breaks during test	16.8	2.1	12.5	0.4	-4.3	-2.0	0.05
Read occasionally in English	16.4	3.1	6.6	0.4	-9.8	-3.2	0.00
Cueing	2.4	0.7	4.1	0.3	1.7	2.3	0.02
School staff administer test	3.0	0.9	2.8	0.2	-0.2	-0.2	0.85
One on one	5.6	1.8	2.6	0.2	-3.0	-1.7	0.09
Preferential seating	7.0	1.8	2.5	0.2	-4.4	-2.4	0.02
Scribe	4.0	1.0	1.7	0.2	-2.3	-2.3	0.02

*Difference with SD calculator accommodation

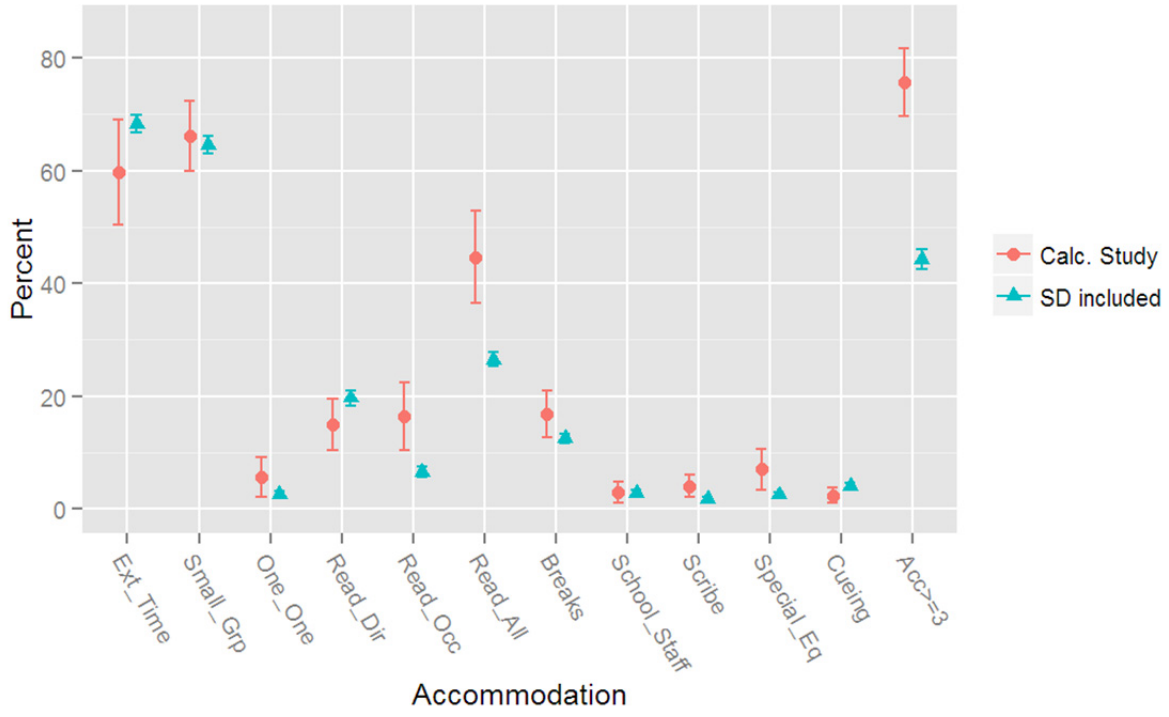


Figure 13. Types and number of accommodations used by grade 8 students participating in NAEP mathematics (2011)

Table 13. Number of accommodations used by grade 8 students participating in NAEP mathematics (2011)

Number of Accommodation	Calculator Study (SD CalcStudy)		Included SD (SD Calc and SD NoCalc)				
	Percent	SE	Percent	SE	Difference*	*t*	*p*-value
Accommodations >= 1	100.0	0.0	83.6	0.5	-16.4	-31.3	0.00
Accommodations >= 2	92.5	2.1	69.6	0.7	-23.0	-10.5	0.00
Accommodations >= 3	75.7	3.0	44.2	0.9	-31.4	-9.9	0.00
Accommodations >= 4	49.1	4.0	14.6	0.5	-34.4	-8.6	0.00
Accommodations >= 5	21.2	2.5	3.6	0.2	-17.6	-7.1	0.00

*Difference with SD calculator accommodation

Comparison of Omit Rates and Items Not Reached

We also examined the percent of items omitted and not reached. Table 14 and Figure 14 present omit rates for students in the calculator study and operational NAEP. Few items were omitted. Students in the calculator study were similar to SD (SD Calc and SD NoCalc) as well as non-SD in operational NAEP in terms of omit rate.

Table 14. Percent of items omitted by grade 8 students participating in NAEP mathematics (2011)

Student Subgroup	Calculator Use		Booklet			Block One			Block Two		
	Block 1	Block 2	Percent Omitted	S.E.	95% C.I.	Percent Omitted	S.E.	95% C.I.	Percent Omitted	S.E.	95% C.I.
Calculator study (SD CalcStudy)	Y	Y	1.8	0.2	[1.5 , 2.1]	1.9	0.2	[1.4 , 2.3]	1.8	0.2	[1.3 , 2.2]
SD with calculator accommodation--operational NAEP (SD Calc)	N	N	1.5	0.2	[1.0 , 1.9]	1.2	0.2	[0.8 , 1.5]	1.8	0.4	[1.0 , 2.6]
	N	Y	3.6	0.9	[1.8 , 5.3]	3.2	1.1	[1.0 , 5.4]	3.9	0.9	[2.1 , 5.8]
	Y	N	2.5	0.7	[1.1 , 3.9]	2.6	0.8	[1.0 , 4.1]	2.4	0.7	[1.0 , 3.9]
	y	y	1.7	0.4	[0.9 , 2.5]	1.7	0.4	[1.0 , 2.5]	1.6	0.6	[0.5 , 2.8]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	1.6	0.2	[1.3 , 2.0]	1.8	0.2	[1.4 , 2.3]	1.4	0.2	[1.1 , 1.8]
Non-SD--operational NAEP (NonSD)	Y	Y	1.9	0.1	[1.7 , 2.1]	1.9	0.1	[1.7 , 2.1]	1.9	0.1	[1.7 , 2.1]

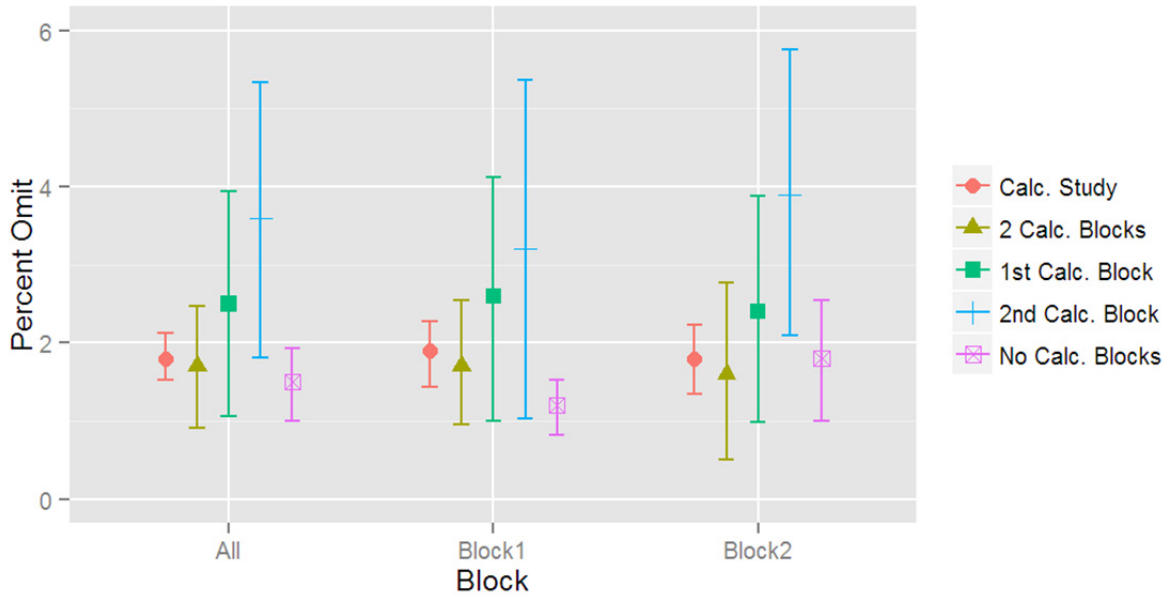


Figure 14. Percent of items omitted by grade 8 students participating in NAEP mathematics (2011)

Percent of items not reached is shown in Table 15 and Figure 15. Students receiving two calculator blocks in operational NAEP (SD and non-SD), completed slightly more items than SD in the calculator study who also received two calculator blocks.

Table 15. Percent of items not reached by grade 8 students participating in NAEP mathematics (2011)

	Calculator Use		Booklet			Block One			Block Two		
Student Subgroup	Block 1	Block 2	Pct. Not Reached	S.E.	95% C.I.	Pct. Not Reached	S.E.	95% C.I.	Pct. Not Reached	S.E.	95% C.I.
Calculator study (SD CalcStudy)	Y	Y	1.4	0.6	[0.2 , 2.6]	1.3	0.6	[0.2 , 2.4]	1.4	0.7	[0.0 , 2.7]
SD with calculator accommodation--operational NAEP (SD Calc)	N	N	2.0	0.7	[0.7 , 3.4]	2.3	0.8	[0.7 , 3.9]	1.7	0.6	[0.5 , 2.9]
	N	Y	3.4	1.3	[0.9 , 5.9]	5.8	2.6	[0.7 , 10.9]	1.1	0.5	[0.1 , 2.2]
	Y	N	1.2	0.5	[0.2 , 2.2]	0.8	0.4	[0.0 , 1.6]	1.6	0.7	[0.2 , 2.9]
	y	y	0.0	0.0	[0.0 , 0.1]	0.0	0.0	[0.0 , 0.1]	0.0	0.0	[0.0 , 0.1]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	0.6	0.2	[0.2 , 1.0]	0.8	0.3	[0.2 , 1.4]	0.5	0.2	[0.2 , 0.8]
Non-SD--operational NAEP (NonSD)	Y	Y	1.1	0.1	[0.9 , 1.3]	1.3	0.1	[1.1 , 1.5]	0.9	0.1	[0.6 , 1.1]

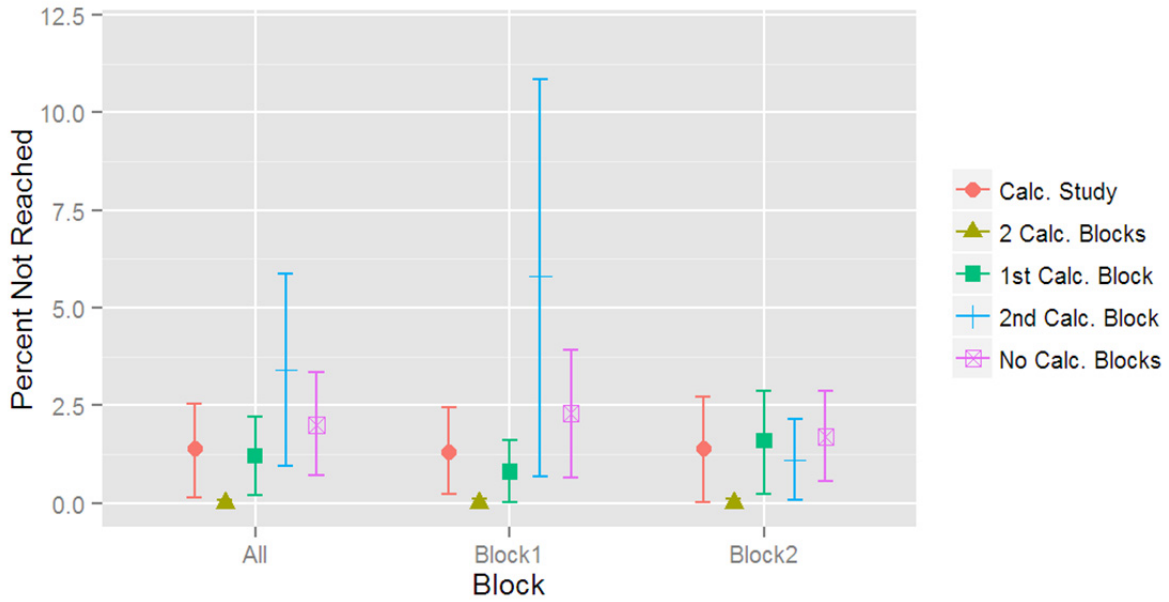


Figure 15. Percent of items not reached by grade 8 students participating in NAEP mathematics (2011)

Comparison of Calculator Item Types

Items in calculator blocks may or may not require a calculator for ease in responding. Items containing larger numbers are typically easier with a calculator and are designated "calculator active." Items for which a calculator would not be helpful, for example, identifying a geometric shape, are considered "calculator inactive." "Calculator neutral" items are those where a calculator could be used but usually wouldn't be needed, such as simple computations.

Calculator active items were more difficult than calculator neutral or inactive items for all students as shown in Table 16 and Figure 16. In general, there was no difference between calculator inactive and neutral items.

Table 16. Percent correct by calculator activity of grade 8 students participating in NAEP mathematics (2011)

	Calculator Use		Calculator Active			Calculator Inactive			Calculator Neutral		
Student Subgroup	Block 1	Block 2	Percent Correct	S.E.	95% C.I.	Percent Correct	S.E.	95% C.I.	Percent Correct	S.E.	95% C.I.
Calculator study (SD CalcStudy)	Y	Y	16.6	1.1	[14.4 , 18.9]	30.8	1.6	[27.7,34.0]	32.9	1.1	[30.8 , 35.0]
SD with calculator accommodation--operational NAEP (SD Calc)	1 or 2 Calc. Blocks		17.1	1.2	[14.7 , 19.6]	31.3	2.0	[27.4,35.1]	28.3	1.2	[25.9 , 30.7]
SD without calculator accommodation--operational NAEP (SD NoCalc)	Y	Y	20.6	1.2	[18.3 , 22.9]	29.9	0.9	[28.1,31.6]	38.4	0.7	[37.1 , 39.7]
Non-SD--operational NAEP (NonSD)	Y	Y	43.4	0.5	[42.5 , 44.3]	51.3	0.4	[50.5,52.1]	50.8	0.4	[50.1 , 51.5]

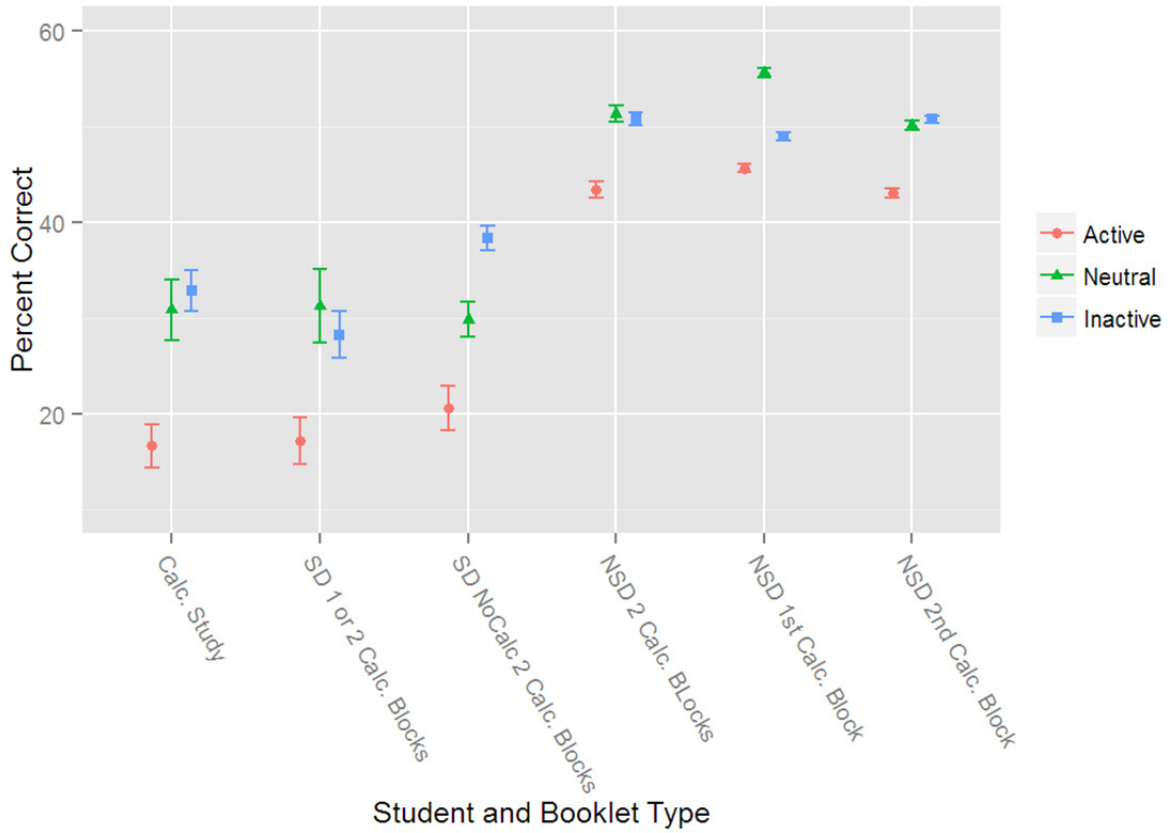


Figure 16. Percent correct by calculator activity of grade 8 students participating in NAEP mathematics (2011)

Discussion

Offering calculator accommodations for NAEP mathematics assessments increases inclusion, especially in states where a calculator accommodation is offered on the state assessment and thus is often included on students' IEPs. The key questions are: (a) how does the increased inclusion affect student performance? (research question 1) and (b) is there an impact on reporting of results, particularly at the state level? (research question 2) In addressing these questions, it is useful to look at the students requiring calculator accommodations and to identify similarities and differences in students with and without such accommodations (research questions 3-5). In 2011, a NAEP calculator booklet special study was conducted offering a calculator accommodation to students who were to be excluded from participating in NAEP mathematics. Using data from the grade 8 special study participants as well as students in the reporting sample, both SD and non-SD, we performed analyses to address the key questions.

Effects on Student Performance

Students in the calculator study achieved mean scale scores similar to other students with disabilities (237.8 compared to 235.7 to 240.9 for SD requiring a calculator accommodation, but participating without one and 246.0 for SD not requiring a calculator accommodation). Both

calculator study and operational NAEP SD participants scored lower than non-SD (who averaged 286.1). Similarly, using overall percent correct reveals the same pattern of SD in the calculator study with 27 percent correct and other SD scoring between 25 percent and 31 percent correct. In comparison, non-SD scored between 48 percent and 50 percent correct, significantly higher than SD, depending on how many calculator blocks they received. Students with disabilities who do not require calculator accommodations tended to perform at a slightly higher level than students in the calculator study. The confidence interval of mean scale scores overlaps for SD without calculator accommodations but who received two calculator blocks and SD in the calculator study (also with two calculator blocks). Comparisons of overall percent correct showed no overlap. The similarity in mean scores for SD tested with and without a calculator accommodation suggests that providing a calculator accommodation for SD, when appropriate, should have little, if any, effect on performance of students with disabilities. However, reducing the proportion of SD excluded from NAEP will lower overall mean score estimates.

Additional measures of student performance included the percentage of items omitted and the percentage of items not reached. Students tended to omit few items; there were no differences in omit rates between those in the calculator study and other SD who required calculator accommodations. NAEP is not considered a speeded test. Thus, students generally finished the assessment, leaving very few items unanswered at the end of each block. Students with disabilities who were included in NAEP operationally without calculator accommodations, but received two calculator blocks, were slightly more likely to finish the assessment than students in the calculator study who also received two calculator blocks. This outcome was evident in the first block but not the second block.

Effects on State Reporting

Increasing inclusion on NAEP mathematics assessments by offering calculator accommodations is largely a state issue, driven in part by state policies of allowing the use of calculators on state assessments. The use of calculators on NAEP is likely to extend beyond states with assessment policies allowing calculators, as students in states that do not allow calculators on state exams may use them in classroom instruction and may have a calculator accommodation on their IEPs. Such students may be eligible for calculator accommodations on NAEP.

The effect on state level reporting is a function of the number of students using the accommodation. The more students that a state includes who were previously excluded from NAEP, the greater the impact on state level reporting. This was clearly seen in Oklahoma which would have had the greatest drop in exclusion rates if they had included students who participated in the calculator study. States exhibiting performance losses on NAEP will likely experience even greater losses or more subgroups losing ground, such as noted in Maryland which would have had a relatively large drop in its exclusion rate with the addition of a calculator accommodation. There were fewer impacts on state reporting in New Jersey and Texas where there would have been smaller decreases in exclusion rates.

States with a large number of students participating in NAEP with calculator accommodations will need to determine the extent of the impact on NAEP results. The calculator accommodation may be an additional accommodation for students who previously would have participated, or it may be the accommodation that allows a previously excluded student to be included. Increases in inclusion, particularly large increases, may affect NAEP state results because there is a performance gap between students with disabilities and those with no disabilities. Increases in the SD population participating in NAEP have the potential to decrease NAEP state results.

States with the largest drop in exclusion rates may need additional assistance to explain changes in NAEP policy when NAEP state data are released.

Comparison of Students Requiring Calculator Accommodations

With respect to student performance, individuals in the calculator study were similar to other students with disabilities, including those who use calculator accommodations on state assessments. This is true for SD who participate in NAEP without calculator accommodations as well as SD who do not need a calculator accommodation. The gap between SD and non-SD remains, with non-SD outscoring their SD peers.

We compared students in the calculator study to those participating in NAEP as well as those excluded from NAEP. The calculator study included more males than females. There was no appreciable difference in the gender distribution of students in the calculator study compared to SD included in NAEP or SD excluded from NAEP. The calculator study included somewhat more minorities than other SD groups (e.g., SD Calc and SD NoCalc), most closely resembling SD excluded from NAEP. Similarly, in the calculator study more students received free or reduced meals than other SD groups, particularly non-SD, and most closely resembling excluded students. Thus, there are some differences between students with disabilities in the calculator study and SD in the reporting sample of NAEP with appropriate accommodations (but not a calculator accommodation). Overall, calculator study students are similar to SD in operational NAEP but more likely to be minorities participating in the NSLP.

The degree or severity of disability as well as the different type of disability lead to differences in grade level performance. Students in the calculator study were generally reported as having mild or moderate disabilities as were SD in the reporting sample. In contrast, students excluded from NAEP were those with the most profound disabilities. The degree of disability is related to the grade level of performance, generally the milder the disability the closer to grade level the student performs, and the more profound the disability the farther from grade level the student performs. Students in the calculator study were slightly more likely than SD included in NAEP to be 2 or more years below grade level and less likely to be on or above grade level. However, students in the calculator study more closely resembled SD included in NAEP than excluded students. The major types of disabilities affecting SD were similar across students in the calculator study and those included in NAEP.

There are some differences in the accommodations needed by SD in the calculator study and those in the reporting sample of NAEP. Students in the calculator study were much more likely to need three or more accommodations. In addition, those in the calculator study were more likely to use “read aloud occasionally” and “read aloud” all than SD in operational NAEP.

Conclusion

Data from the 2011 calculator study indicate that students who were previously excluded from NAEP because calculator accommodations were not available may reasonably participate in NAEP. Performance of students requiring calculator accommodations is not significantly different from other SD, but is lower than students without disabilities. The gap between SD and non-SD contributes to the size of the effect on state-level reporting. The more students who were excluded on the basis of calculator accommodations in a state, the greater the impact on state results.

Although there are some similarities and some differences between students in the calculator study (i.e., those who were previously excluded but may be included with a calculator accommodation),

the differences are small and don't have a significant impact on student performance or participation. Students with profound disabilities and students performing well below grade level would still be excluded even with the availability of a calculator accommodation on NAEP.

References

Accommodations and Inclusion Coordinating Council (AICC). (May 24, 2012). *Teleconference – discussion of results and conclusions from the 2011 calculator booklet special study*.

Educational Testing Service (ETS). (March 26, 2012). Memorandum from Meng Wu and Dave Freund (ETS), to Pat Etienne (NCES). *Summary of results and conclusions from the 2011 Inclusion study.* Princeton, NJ: Educational Testing Service.

Fletcher, J.M.,Francis, D.J., O'Malley, K., Copeland, K., Mehta, P., Caldwell, C.J., … Vaughn, S. (2009). Effects of a bundled accommodations package on high-stakes testing for middle school students with reading disabilities. *Exceptional Children, 75*(4), 447-463. Retrieved March 26, 2013 from http://www.freepatnetsonline.com/article/Exceptional-Children/202627457.html.

Freund, D. (2012, May). Follow up to May 1, 2012 AICC meeting. Informal written communication. Princeton, NJ: Educational Testing Service.

Frieden, L. (2004). Improving educational outcomes for students with disabilities. Washington, DC: National Council on Disability.

Haertel, E.H. (2003, December). *Including Students with Disabilities and English Language Learners in NAEP: Effects of Differential Inclusion Rates on Accuracy and Interpretability of Findings*. Unpublished paper presented at NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP, Retrieved May 3, 2012 from http://www.eric.ed.gov/PDFS/ED500430.pdf

National Assessment Governing Board (NAGB). (2007). *Assessment and Item Specifications for the NAEP 2009 Mathematics* Assessment. Washington, DC: U.S. Department of Education.

National Assessment Governing Board (NAGB). (2010). *NAEP Testing and Reporting on Students with Disabilities and English Language Learners: Policy Statement*. Washington, DC: U.S. Department of Education. Retrieved May 3, 2012, from http://www.nagb.org/policies/PoliciesPDFs/Reporting%20and%20Dissemination/naep_testandreport_studentswithdisabilities.pdf

National Center for Education Statistics (NCES). (2011). *The Nation's Report Card: Mathematics 2011* (NCES 2012–458). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved May 8, 2012, from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012458.pdf

National Center for Education Statistics (NCES). Embargoed data.

Comparison of change in average reading scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 4: By state

Jurisdiction	Reported 2013	Reported 2011	Reported 2013-2011	FPE 2013	FPE 2011	FPE 2013-2011	Reported-FPE difference
Alabama	218.6	220.3	-1.7	217.8	218.7	-0.9	-0.8
Alaska	209.3	207.9	1.4	208.4	206.8	1.6	-0.2
Arizona	213.1	212.4	0.7	212.4	211.3	1.1	-0.3
Arkansas	218.5	216.5	2.0	217.8	215.7	2.2	-0.1
California	212.5	211.4	1.2	210.7	209.9	0.8	0.3
Colorado[1]	226.7	223.4	3.2 *	225.6	222.4	3.2	0.1
Connecticut	229.6	227.4	2.2	228.7	226.0	2.6	-0.5
Delaware	225.8	225.1	0.6	222.8	221.0	1.8	-1.1
District of Columbia	205.6	200.6	5.0 *	204.8	199.0	5.8 *	-0.8
DoDEA[2]	232.1	229.5	2.6 *	229.0	226.0	3.0 *	-0.4
Florida	227.5	224.5	2.9	226.0	223.3	2.7	0.3
Georgia	221.8	220.8	1.0	219.2	217.2	2.0	-1.0
Hawaii	214.8	213.6	1.2	213.6	212.4	1.2	0.0
Idaho	219.3	220.8	-1.5	218.2	219.4	-1.1	-0.4
Illinois	218.5	219.4	-0.8	217.9	218.3	-0.3	-0.5
Indiana	225.3	220.7	4.6 *	224.1	220.0	4.1 *	0.5
Iowa	223.8	220.7	3.1 *	223.1	219.9	3.1 *	0.0
Kansas	223.4	223.6	-0.2	222.3	222.0	0.3	-0.4
Kentucky	224.4	225.1	-0.7	222.5	220.2	2.3	-3.0
Louisiana	210.5	210.4	0.0	209.8	209.7	0.2	-0.1
Maine	224.8	222.0	2.8 *	223.9	221.3	2.7 *	0.1
Maryland	232.1	230.8	1.3	224.5	225.6	-1.1	2.3
Massachusetts	232.4	236.8	-4.4 *	230.8	233.9	-3.1 *	-1.3
Michigan	217.4	218.9	-1.5	215.0	216.4	-1.4	-0.1
Minnesota	227.0	222.3	4.7 *	225.4	221.2	4.2 *	0.6
Mississippi	208.5	209.2	-0.7	208.2	208.6	-0.4	-0.3
Missouri	222.3	220.3	2.0	221.6	219.3	2.2	-0.3
Montana	223.0	225.1	-2.2 *	221.2	222.4	-1.2	-1.0
Nebraska	223.3	223.3	-0.1	221.0	220.7	0.3	-0.4
Nevada	213.8	212.6	1.2	212.9	211.9	1.0	0.2
New Hampshire	232.0	230.4	1.6	230.7	229.1	1.6	0.0
New Jersey	228.7	231.2	-2.5	227.8	227.1	0.7	-3.2
New Mexico	205.8	208.0	-2.3	205.2	205.1	0.1	-2.3
New York	223.8	222.5	1.3	223.1	221.3	1.7	-0.4
North Carolina	222.2	221.4	0.9	221.0	219.9	1.2	-0.3
North Dakota[1]	224.1	225.6	-1.5 *	221.5	222.1	-0.6	-0.9
Ohio	223.9	223.8	0.1	222.5	220.5	2.0	-1.9
Oklahoma[1]	217.0	215.5	1.5	215.9	212.4	3.4 *	-1.9
Oregon	219.1	216.4	2.7	217.2	214.4	2.9	-0.2
Pennsylvania	226.4	227.2	-0.8	225.1	225.2	-0.1	-0.8
Rhode Island	222.8	222.5	0.3	221.9	221.3	0.6	-0.3
South Carolina	213.6	214.9	-1.3	212.4	212.8	-0.5	-0.8
South Dakota	217.9	219.8	-1.9	216.1	217.8	-1.6	-0.2
Tennessee	219.7	214.6	5.1 *	217.8	210.3	7.5 *	-2.4
Texas	216.9	218.3	-1.4	214.1	213.2	1.0	-2.4
Utah	222.8	220.4	2.4	220.6	217.7	2.9	-0.5
Vermont[1]	228.0	226.8	1.2	227.3	225.3	1.9 *	-0.7
Virginia	228.6	226.4	2.2	227.4	224.4	3.1	-0.9
Washington	225.0	220.5	4.5 *	223.1	218.7	4.4 *	0.1
West Virginia	214.7	214.4	0.3	213.5	213.1	0.4	-0.1

Wisconsin	220.8	221.2	-0.4	219.7	219.9	-0.3	-0.2
Wyoming[1]	225.8	224.1	1.8	225.1	222.9	2.1 *	-0.4

* Statistically significant change.

[1]The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.

[2]Department of Defense Education Activity (domestic and overseas schools).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 and 2013 Reading Assessments.

Read more about research on exclusion rates.

Last updated 05 November 2013 (FW)

Comparison of change in average mathematics scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 4: By state

Jurisdiction	Reported 2013	Reported 2011	Reported 2013-2011	FPE 2013	FPE 2011	FPE 2013-2011	Reported-FPE difference
Alabama	232.9	231.3	1.6	232.3	230.6	1.6	0.0
Alaska	236.1	236.4	-0.2	235.7	235.5	0.2	-0.5
Arizona	240.3	235.2	5.1 *	239.7	234.6	5.1 *	0.1
Arkansas	239.9	237.8	2.1	239.3	237.4	1.9	0.2
California	233.7	234.2	-0.5	232.6	233.4	-0.8	0.3
Colorado[1]	247.0	244.5	2.5 *	246.3	243.9	2.4	0.1
Connecticut	243.4	242.4	1.0	242.8	241.9	0.9	0.2
Delaware	243.1	240.4	2.7 *	242.1	239.0	3.1 *	-0.4
District of Columbia	228.6	221.8	6.8 *	228.0	219.8	8.2 *	-1.5
DoDEA[2]	245.0	240.8	4.2 *	244.3	239.9	4.4 *	-0.2
Florida	241.7	239.8	1.8	240.9	239.2	1.7	0.1
Georgia	240.0	238.4	1.7	239.3	237.6	1.7	0.0
Hawaii	243.3	238.8	4.5 *	242.6	237.8	4.8 *	-0.3
Idaho	240.7	240.3	0.4	240.0	239.7	0.3	0.1
Illinois	239.0	238.8	0.2	238.6	237.8	0.8	-0.6
Indiana	248.6	243.8	4.8 *	248.0	242.9	5.1 *	-0.3
Iowa	245.8	242.6	3.2 *	245.4	242.0	3.4 *	-0.2
Kansas	246.2	246.3	-0.1	245.5	245.5	0.0	-0.1
Kentucky	241.5	240.8	0.6	240.8	239.7	1.1	-0.5
Louisiana	231.4	230.8	0.6	230.9	230.3	0.6	0.0
Maine	245.8	244.3	1.6	244.8	243.6	1.2	0.3
Maryland	245.2	247.1	-1.9	244.7	245.2	-0.6	-1.4
Massachusetts	253.0	253.4	-0.4	252.1	252.2	-0.1	-0.2
Michigan	236.8	236.4	0.4	236.0	235.4	0.6	-0.2
Minnesota	253.4	249.2	4.2 *	252.8	248.6	4.2 *	0.0
Mississippi	231.1	229.9	1.2	230.8	229.5	1.2	0.0
Missouri	239.5	240.5	-0.9	238.8	239.8	-1.0	0.0
Montana	243.7	243.8	-0.1	242.7	243.0	-0.3	0.2
Nebraska	243.2	239.8	3.4 *	242.4	239.2	3.2 *	0.1
Nevada	236.3	237.0	-0.8	235.7	236.2	-0.5	-0.2
New Hampshire	253.0	251.8	1.2	252.5	251.1	1.4	-0.2
New Jersey	246.9	248.0	-1.1	246.3	246.9	-0.5	-0.6
New Mexico	232.8	232.8	-0.1	232.3	232.1	0.2	-0.3
New York	240.3	237.5	2.8 *	239.8	237.1	2.8 *	0.1
North Carolina	244.8	244.5	0.3	244.2	243.7	0.5	-0.2
North Dakota	246.4	245.2	1.3 *	245.1	243.6	1.5 *	-0.2
Ohio	245.5	244.0	1.5	244.9	243.0	1.9	-0.4
Oklahoma[1]	238.9	237.4	1.5	238.1	234.1	4.0 *	-2.5
Oregon[1]	240.1	236.9	3.2	239.1	235.6	3.5 *	-0.3
Pennsylvania	244.0	245.7	-1.6	243.2	244.9	-1.7	0.1
Rhode Island	241.4	241.6	-0.2	240.9	241.3	-0.4	0.2
South Carolina	236.6	237.3	-0.7	236.0	236.7	-0.6	-0.1
South Dakota	241.0	241.0	0.0	240.2	240.0	0.2	-0.2
Tennessee	239.8	232.9	6.9 *	239.1	231.5	7.6 *	-0.7
Texas	241.9	241.1	0.8	241.1	239.5	1.6	-0.8
Utah	242.8	242.5	0.3	242.2	241.8	0.4	-0.1
Vermont	247.8	246.6	1.2	247.1	246.0	1.1	0.1
Virginia	246.2	245.3	0.8	245.5	244.4	1.1	-0.2
Washington	246.3	243.2	3.1 *	245.2	242.4	2.9 *	0.2
West Virginia	237.4	234.7	2.8 *	236.7	234.0	2.7 *	0.1

Wisconsin	244.7	244.7	0.0	243.8	243.9	-0.1	0.2
Wyoming	246.5	243.9	2.6 *	246.1	243.3	2.8 *	-0.2

* Statistically significant change.

[1]The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.

[2]Department of Defense Education Activity (domestic and overseas schools).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 and 2013 Mathematics Assessments.

Read more about research on exclusion rates.

Last updated 05 November 2013 (FW)

Comparison of change in average reading scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 8: By state

	Reported			FPE			Reported-FPE
Jurisdiction	2013	2011	2013-2011	2013	2011	2013-2011	difference
Alabama	257.4	258.4	-1.0	256.6	257.1	-0.5	-0.5
Alaska	261.3	261.3	0.0	260.6	260.4	0.2	-0.2
Arizona	260.4	260.1	0.3	259.5	259.5	0.0	0.3
Arkansas[1]	262.0	259.1	2.8 *	260.7	258.2	2.6	0.3
California	261.5	254.9	6.6 *	260.1	253.6	6.5 *	0.1
Colorado	271.0	270.6	0.4	270.3	269.6	0.8	-0.4
Connecticut	274.5	274.7	-0.2	273.2	273.4	-0.2	0.0
Delaware	266.0	265.8	0.1	264.2	263.2	0.9	-0.8
District of Columbia	247.7	242.1	5.7 *	246.8	240.8	6.0 *	-0.3
DoDEA[2]	277.0	272.3	4.6 *	275.0	270.6	4.4 *	0.2
Florida	265.8	262.1	3.7 *	264.9	260.8	4.1 *	-0.3
Georgia	264.6	262.4	2.3	262.5	260.3	2.2	0.1
Hawaii	260.0	257.2	2.8 *	258.9	255.9	3.0 *	-0.2
Idaho	270.2	268.0	2.2	269.1	267.0	2.1	0.1
Illinois	266.9	265.6	1.3	266.2	264.7	1.5	-0.2
Indiana	267.3	264.7	2.5	266.2	263.4	2.8	-0.3
Iowa	269.0	264.6	4.4 *	268.3	264.1	4.2 *	0.3
Kansas	266.9	267.3	-0.4	265.9	266.2	-0.3	-0.2
Kentucky[1]	269.6	268.8	0.8	267.8	265.2	2.5 *	-1.8
Louisiana	257.4	254.7	2.7	256.8	254.3	2.5	0.1
Maine	269.2	269.9	-0.7	268.5	269.0	-0.5	-0.1
Maryland	273.8	271.2	2.6	269.0	266.9	2.1	0.5
Massachusetts[1]	277.0	275.4	1.6	275.7	272.3	3.4 *	-1.8
Michigan	265.9	265.2	0.7	264.2	262.3	2.0	-1.2
Minnesota	271.0	270.1	0.9	269.7	268.4	1.3	-0.4
Mississippi	253.1	253.8	-0.7	252.8	253.2	-0.4	-0.3
Missouri	267.2	266.8	0.5	266.6	265.8	0.8	-0.4
Montana	271.8	272.9	-1.1	270.5	270.6	-0.1	-1.0
Nebraska[1]	269.2	267.7	1.5	267.6	264.9	2.7 *	-1.2
Nevada	261.7	258.2	3.5 *	261.1	257.2	4.0 *	-0.5
New Hampshire	274.3	272.1	2.2 *	272.9	270.2	2.7 *	-0.5
New Jersey	276.4	275.2	1.2	275.3	272.2	3.1	-1.9
New Mexico	255.9	255.9	0.0	255.1	253.3	1.7	-1.7
New York	266.3	265.7	0.6	265.8	264.3	1.5	-0.9
North Carolina	264.5	262.9	1.6	263.5	261.8	1.8	-0.2
North Dakota	267.8	268.7	-0.9	265.2	264.9	0.3	-1.2
Ohio	269.1	268.3	0.8	267.8	265.5	2.4	-1.6
Oklahoma[1]	261.9	260.1	1.8	261.1	257.8	3.3 *	-1.5
Oregon	268.3	264.2	4.1 *	267.4	262.9	4.5 *	-0.5
Pennsylvania	272.1	267.8	4.3 *	271.1	266.1	5.0 *	-0.6
Rhode Island	266.7	265.1	1.6	265.9	264.4	1.5	0.1
South Carolina	261.4	260.3	1.1	260.3	257.6	2.7	-1.6
South Dakota	268.1	268.9	-0.8	266.4	266.9	-0.5	-0.3
Tennessee	265.4	259.2	6.2 *	263.7	255.8	7.9 *	-1.7
Texas[1]	263.7	261.4	2.3	261.7	257.7	4.0 *	-1.8
Utah	270.0	267.1	3.0 *	268.4	264.8	3.5 *	-0.6
Vermont	274.4	273.8	0.5	273.9	272.2	1.7	-1.2
Virginia	267.6	267.3	0.3	266.9	265.3	1.6	-1.3
Washington	272.0	267.6	4.5 *	270.6	266.2	4.4 *	0.1
West Virginia	257.4	256.1	1.3	256.3	255.1	1.2	0.1

Wisconsin	267.5	267.2	0.4	266.6	265.8	0.8	-0.4
Wyoming	271.0	269.6	1.4	270.5	268.5	2.0	-0.6

* Statistically significant change.

[1]The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.

[2]Department of Defense Education Activity (domestic and overseas schools).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 and 2013 Reading Assessments.

Read more about research on exclusion rates.

Last updated 05 November 2013 (FW)

Comparison of change in average mathematics scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 8: By state

Jurisdiction	Reported 2013	Reported 2011	Reported 2013-2011	FPE 2013	FPE 2011	FPE 2013-2011	Reported-FPE difference
Alabama	269.2	269.1	0.1	268.4	268.3	0.1	0.0
Alaska	281.6	283.3	-1.7	281.0	281.8	-0.8	-0.9
Arizona	279.7	279.0	0.7	278.9	278.4	0.5	0.2
Arkansas	277.9	279.1	-1.2	276.6	278.2	-1.6	0.4
California	275.9	272.8	3.1	275.1	272.1	2.9	0.2
Colorado	289.7	291.7	-2.1	288.8	291.2	-2.3	0.3
Connecticut	285.2	287.0	-1.8	284.0	286.2	-2.3	0.5
Delaware	282.3	282.8	-0.4	281.7	281.2	0.5	-0.9
District of Columbia	265.3	260.5	4.8 *	264.8	258.6	6.2 *	-1.4
DoDEA[2]	290.4	287.9	2.6 *	289.8	286.4	3.3 *	-0.8
Florida	280.9	277.8	3.0 *	279.9	276.9	3.0 *	0.1
Georgia	279.2	278.5	0.7	278.2	277.0	1.2	-0.5
Hawaii	281.4	277.8	3.6 *	280.5	276.7	3.8 *	-0.2
Idaho	286.4	286.7	-0.3	285.7	285.9	-0.2	-0.1
Illinois	284.9	283.2	1.7	284.3	282.0	2.3	-0.6
Indiana	287.8	285.0	2.8	286.8	283.9	2.9	-0.2
Iowa	285.1	284.9	0.1	284.5	284.1	0.4	-0.3
Kansas	289.5	289.6	-0.1	288.4	288.9	-0.5	0.4
Kentucky	280.6	281.6	-1.0	279.3	280.2	-0.9	-0.1
Louisiana	272.8	272.8	-0.1	272.1	272.2	-0.1	0.0
Maine	288.7	288.7	0.0	288.0	287.9	0.1	-0.1
Maryland	286.6	288.0	-1.4	285.7	284.5	1.2	-2.5
Massachusetts[1]	300.6	298.5	2.1	299.4	296.5	2.9 *	-0.9
Michigan	280.1	280.2	0.0	278.8	278.2	0.6	-0.6
Minnesota	294.6	294.9	-0.4	293.4	293.7	-0.3	-0.1
Mississippi	271.2	269.2	1.9	270.7	268.6	2.1	-0.2
Missouri	283.0	281.9	1.0	282.2	281.2	1.0	0.0
Montana	289.2	292.9	-3.7 *	288.3	291.8	-3.5 *	-0.1
Nebraska[1]	285.1	283.2	1.9	283.9	281.2	2.7 *	-0.8
Nevada	278.3	278.1	0.2	277.7	276.6	1.1	-0.9
New Hampshire	295.7	292.1	3.6 *	295.1	291.2	3.9 *	-0.3
New Jersey	296.1	294.1	1.9	295.0	292.4	2.5	-0.6
New Mexico	272.8	274.5	-1.7	271.8	273.6	-1.8	0.1
New York	281.8	280.5	1.4	280.9	279.9	1.0	0.3
North Carolina	285.6	286.3	-0.6	284.8	285.2	-0.4	-0.2
North Dakota	290.5	292.0	-1.5	288.7	290.0	-1.3	-0.2
Ohio	289.5	288.6	0.9	288.6	286.0	2.6	-1.7
Oklahoma[1]	275.5	279.2	-3.7 *	274.5	274.0	0.6	-4.2
Oregon	283.5	282.5	1.0	282.6	281.7	0.9	0.1
Pennsylvania	289.6	286.1	3.5 *	288.7	284.6	4.1 *	-0.6
Rhode Island	284.1	282.9	1.2	283.5	282.2	1.3	-0.1
South Carolina	279.8	281.0	-1.2	279.0	278.9	0.1	-1.2
South Dakota	287.3	290.6	-3.3 *	286.4	289.4	-3.0 *	-0.3
Tennessee	277.7	274.0	3.7 *	276.6	272.1	4.5 *	-0.8
Texas	288.2	290.3	-2.1	287.1	287.9	-0.8	-1.4
Utah	284.3	283.3	1.0	283.3	281.7	1.5	-0.5
Vermont	295.5	293.9	1.6	294.9	293.2	1.7	-0.1
Virginia	288.1	289.3	-1.1	287.5	287.5	0.0	-1.1
Washington	290.0	288.1	1.8	288.7	287.1	1.6	0.3
West Virginia	274.4	273.3	1.2	273.3	272.4	0.9	0.3

Wisconsin	288.7	288.7	0.1	287.7	287.4	0.3	-0.2
Wyoming	288.1	287.8	0.4	287.3	287.2	0.2	0.2

* Statistically significant change.
[1]The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.
[2]Department of Defense Education Activity (domestic and overseas schools).
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 and 2013 Mathematics Assessments.

Read more about research on exclusion rates.

Last updated 05 November 2013 (FW)

A Validity Study of the NAEP Full Population Estimates

Larry V. Hedges
Northwestern University

Victor Bandeira de Mello
American Institutes for Research

August 2013
Commissioned by the NAEP Validity Studies (NVS) Panel

EXCERPTS

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

This report was prepared for the National Center for Education Statistics under Contract No. ED-04-CO-0025/0012 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

2047_08/13

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Peter Behuniak
University of Connecticut

George W. Bohrnstedt
American Institutes for Research

James R. Chromy
Research Triangle Institute

Phil Daro
Strategic Education Research Partnership (SERP) Institute

Lizanne DeStefano
University of Illinois

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Larry Hedges
Northwestern University

Gerunda Hughes
Howard University

Robert Linn
University of Colorado Boulder

Ina V.S. Mullis
Boston College

Scott Norton
Council of Chief State School Officers

Gary Phillips
American Institutes for Research

Lorrie Shepard
University of Colorado Boulder

David Thissen
University of North Carolina, Chapel Hill

Karen Wixson
University of North Carolina, Greensboro

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Janis Brown
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403
Phone: 650/ 843-8100
Fax: 650/ 843-8200

Executive Summary

To support an internal evaluation of the impact of changing exclusion rates on reports of statistically significant gains in National Assessment of Educational Progress (NAEP) scores across states, the National Center for Education Statistics (NCES) sponsored research on imputation procedures used to calculate NAEP scores for the excluded students and provided adjusted or full population estimates (FPEs) for the 1996 to 2000 NAEP mathematics gains. The FPE methodology developed by McLaughlin (2005) makes use of information in the student-level NAEP data file, which includes data for students with disabilities (SDs) and English language learners (ELLs) generated from questionnaire responses completed by school staff. In 2009, the task force on FPEs formed by the National Institute of Statistical Sciences and the NAEP-Education Statistics Services Institute (NAEP-ESSI) found that methods used to calculate FPEs were sufficiently sound that there was no identified need for drastic modifications. The task force also recommended that NCES support studies to extend and further validate the methodology for imputing plausible values. The occasion of two special inclusion studies conducted in conjunction with the 2011 NAEP Mathematics Assessment presented just such an opportunity for additional validity research.

Both studies focused on the assessment of otherwise-excluded students by offering accommodations that are not allowed in operational NAEP. One study allowed the use of calculators as an accommodation (in states that permit this accommodation on their state assessments). The other provided students with an inclusion booklet made up of Knowledge and Skill Appropriate (KaSA) blocks that were somewhat easier than standard NAEP blocks. In some states, there were students included in both studies (that is, some students included because of the calculator accommodation and other students included because of the inclusion blocks). In other states, only the inclusion block was offered because the states do not allow a calculator accommodation on their state assessments. After school personnel had finalized their exclusion decisions for the operational assessment, they were asked to reconsider whether excluded students could participate using the calculator or KaSA blocks. If they agreed, these students became participants in the special studies. The data from the special studies were scaled with the data from the operational NAEP assessment, and plausible values were created for the participants in the special studies.

Because these 2011 special inclusion studies yielded a sample of students excluded from operational NAEP for whom both NAEP scaled plausible values and FPEs were available, they provided an opportunity to conduct a validity study of the FPEs. The logic was to compare results from an assessment that included the actual scaled scores for some otherwise excluded students (those who could be included with the special accommodations) with results based on the FPEs.

The total number of operationally excluded students in the 2011 NAEP Mathematics Assessment was 5,049 out of a total sample of 169,452 public school

students (about 3.0 percent). Only 1,197 (23.4 percent of the excluded students) participated in the validity study (891 in the special calculator booklet study and 307 in the inclusion booklet study). This was a much smaller sample size than had been expected. Moreover, the special studies sample differed somewhat from the group of excluded students as a whole in ways that are likely to be related to performance on the assessment. In particular, the students in the special studies sample were rated by school personnel as tending to be among the more able of the excluded student group.

Because of the small sample sizes, the differences between the means of the FPEs and of estimates based on scaled plausible values for the otherwise excluded students (overall and for 14 subgroups) resulted in only one significant difference. However, when 95 percent confidence intervals were constructed to examine for possible bias, the resulting intervals ran from 0 to 10 NAEP points, suggesting that the FPEs may tend to overestimate the actual population parameter. This overestimation is not surprising (and indeed was hypothesized to be the case) because the achievement information on which the FPEs are based is only from assessed students.

It is not clear that FPEs have to be unbiased to be useful, however. Unbiased estimation of unobservable assessment scores is probably an impossible goal in any event. *A principled method that leads to smaller bias in estimating a group that is undercovered in a population may be highly desirable.* Excluding a population subgroup because it cannot be assessed is roughly equivalent (for estimating population averages) to imputing the mean of the assessed population. The special studies samples investigated here scored, on the average, at about the 10th percentile of the assessed population. If we interpret the difference between the average FPEs and scaled plausible values from the special studies as bias, then the results presented here suggest that the bias in imputing the mean of the assessed population is approximately 10 times as large as that in using the FPEs.

When one considers the possibility of improving NAEP population estimates by expanding the pool of tested students, the study also offers some insights. First, because of the small numbers of students successfully recruited into the special studies (and the characteristics of these students, who tended to be rated by their schools as among the most able of the excluded students), the studies suggest that offering the calculator block and KaSA booklet accommodations, by themselves, would not have a substantial impact on national parameter estimates. However, results for the FPE estimates on the *entire* excluded population do show nonnegligible impacts on national parameter estimates. This suggests that if accommodations to include more of the currently excluded students could be found, such accommodations could have a nonnegligible impact on national parameter estimates.

Finally, one can question whether the concept of *full* population estimates is sensible. The reason is that the concept of full population estimates presupposes that there is (at least in theory) an assessment score for every student, including those who are currently excluded from the assessment. If there are students

whom we could not conceive of as participating in the assessment under any conditions, then the concept of "the assessment score they would have obtained if they had participated" may not make sense. One might therefore argue that a group that could never be assessed should be excluded from the definition of the population used to draw inferences. By redefining the population, efforts could focus on developing methods to include as many members of the (newly defined) population as possible in operational assessments and on developing methods to impute scores for those excluded.

Validity Considerations of the Validity Study

One threat to what might be called the statistical validity of the study is that the sample size may not be large enough to provide adequate statistical power or precision for the estimates compared. To evaluate this, it is important, as a first step, to examine the sample sizes obtained and determine if the estimates for the operationally excluded subgroup are precise enough to draw conclusions. The logical framework of this study is that of an equivalence (not superiority) study. That is, we conclude that FPEs are valid if the estimates based on them do not differ from those based on scaled plausible values for students in the inclusion samples. Consequently, we must set the smallest difference that is meaningful and determine whether the sample size will yield adequate statistical power to detect such differences. As a guideline, we suggest using the convention of 80 percent power at the 5 percent significance level.

A crude precision analysis can be done by computing the standard error of the difference between the estimate of a population parameter (e.g., the mean) based on full population methods Y_{FPE} and the same estimate based on scaled plausible values Y_{SPV}

$$S = \sqrt{S_{FPE}^2 + S_{SPV}^2 - 2S_{FPE}S_{SPV}r},$$

where S^2_{FPE} and S^2_{SPV} are the variances of Y_{FPE} and Y_{SPV}, and r is an estimate of the correlation between them. A crude estimate of the power to detect a true difference between Y_{FPE} and Y_{INC} of size δ is

$$p = 1 - \Phi(c - \delta/S) + \Phi(-c - \delta/S),$$

where c is the appropriate critical value of the standard normal distribution and $\Phi(x)$ is the standard normal cumulative distribution function.

The question of how large the difference δ must be to be meaningful is more difficult. We used the approach of studying how large δ must be to produce a consequential difference in assessment scores, as in Table 7. Using the standard errors from Table 6, we evaluate the power to detect the smallest bias that would lead to a change in national or subgroup means by 0.5 and 1.0 NAEP scale-score points in Table 10. For the nation and all of the subgroups considered, the power to detect a bias large enough to change the average estimate by 1.0 NAEP scale-score points is essentially 1.0; thus, these studies appear adequately powered to detect biases large enough to produce a change of 1.0 NAEP scale-score point.

The situation is somewhat different with respect to biases large enough to produce a change of 0.5 NAEP scale-score points. In the black and Hispanic subgroups, the power to detect such a change is only about 70 percent. Thus, the special studies cannot be considered definitive in ruling out such biases in the black and Hispanic reporting subgroups. Note also that, although the point estimates of bias for these two subgroups were less than 5 NAEP scale-score points, the upper ends of the

95 percent confidence intervals for the bias estimates in these two groups (17.2 and 20.8, respectively) do exceed the threshold for bias that could cause a 0.5 NAEP scale-score change in national estimates for each of those groups. In other words, the power of these validity studies is not high enough to rule out biases that could change national estimates of the mean in the black and Hispanic reporting subgroups by as much as 0.5 NAEP scale-score points.

Table 10. Power to Detect a Bias in FPEs That Could Produce a Change in Overall Averages of 0.5 or 1.0 Points in Various Groups

	To Detect Overall Bias of 0.5		To Detect Overall Bias of 1.0	
Group	**FPE Bias**	**Power**	**FPE Bias**	**Power**
Nation	19	1.00	38	1.00
Gender				
Male	16	0.98	31	1.00
Female	26	1.00	52	1.00
Race/Ethnicity				
White	22	1.00	44	1.00
Black	13	0.71	26	1.00
Hispanic	18	0.70	37	1.00
Region				
Northeast	18	0.95	37	1.00
Southeast	22	0.99	44	1.00
Central	18	0.99	35	1.00
West	19	0.98	38	1.00

Note: These computations assume a two-sided 5 percent nonsimultaneous significance test.

There also are two threats to internal validity that can be characterized as selection threats. If the school personnel making exclusion decisions know that this is part of a special study, then biases might arise because of experimenter demand characteristics (see Orne, 1962) or Hawthorne effects (Mayo, 1949).[6] We believe that the data collection plan did *not* explicitly characterize this as part of a special study, which should minimize that validity threat.

The second selection threat is that the school personnel might be motivated to exclude from the operational assessment the students that they believe will perform most poorly. Because they are told that the initially excluded students will not be part of the operational assessment, they have no incentive to exclude students they believe will perform most poorly from the validity study. However, any tendency to exclude students whom they believe will perform most poorly from the *operational*

[6] Experimenter effects refer to experimental results that are biased as a result of the study participants' desire to please the researcher. Hawthorne effects are similar. In a classic study of worker productivity at the Western Electric Hawthorne factory, it was shown that the results were less due to the interventions that were put in place than the fact that the workers were being studied, which seemed to increase productivity in and of itself.

assessment could mean that the validity study sample may include students who could have been included in the operational assessment, but who were systematically excluded because they were expected to have poorer performance than those included.

The implications for performance of the sample of operationally excluded students who are in the validity study are unclear because these two factors work in opposite directions. Assuming that, in general, students who *could properly* participate in the operational assessment will perform better than those who could not, adding these excluded students to the validity study sample might artifactually elevate the performance of the students in the validity study. However, if school personnel are correct that the operationally excluded students perform more poorly than included students, they may also perform more poorly than the properly excluded students, which would artifactually reduce the performance of the students in the validity study sample.

The basic validity question is whether the excluded students who participated in the special studies differ from other excluded students in unobservable (or at least unmeasured) ways that are correlated with achievement (holding constant the observables used in creating the FPEs). The fact that the results in Table 6 suggest that estimates of average achievement based on scaled plausible values are slightly smaller than those based on FPEs suggests that this may be the case.

More elaborate statistical modeling to estimate the expected performance of the excluded students also would be possible. For example, suppose that the excluded students are modeled to be the lower tail (the lowest x% of the distribution, where x is the exclusion rate) of the achievement distribution. We could use the assumption of a distribution shape (e.g., normal) to obtain the expected average (and even standard deviation) of the excluded group. Such an analysis would not, however, resolve whether the poorer performance of excluded students was a consequence of proper exclusion (which is consistent with excluded students having poorer performance) or improper exclusion (excluding students who could have participated but who were excluded because they were expected to have poorer performance).

Conclusions

The special inclusions studied here are disappointing in that they made it possible to include in the assessment only about a quarter of the excluded students and less than 1 percent of the total sample. Moreover, the students they made it possible to include appear to be among the most able of the excluded students—those who were "almost able" to be included without the special accommodations. The cost of these special accommodations seems relatively large for the potential benefit achieved.

In general, it appears that the FPEs may tend to overestimate the results based on scaled plausible values in the special studies, although these differences are far from statistically significant. This is not surprising (and indeed was hypothesized to be the case) because the achievement information on which the FPEs are based is from assessed students. Presumably, there are reasons that students are not assessed, and

not all of these depend on observable (or at least observed) characteristics. Thus, any assessed student whose observed characteristics are equivalent to a student who is not assessed differs on some characteristics that are not observed. If this is so, and if these unobserved characteristics are correlated with (also unobserved) assessment scores, then the FPEs would be biased estimates of the assessment scores. More specifically, it is plausible that the unobserved information leading to exclusion is negatively related to assessment scores. If so, then FPEs would overestimate the performance of excluded students.

It is not clear that FPEs have to be unbiased to be useful, however. Unbiased estimation of unobservable assessment scores is probably an impossible goal in any event. *A principled method that leads to smaller bias in estimating a group that is undercovered in a population may be highly desirable.* Excluding a population subgroup because it cannot be assessed is roughly equivalent (for estimating population averages) to imputing the mean of the assessed population. The special studies sample investigated here scored, on the average, at about the 10th percentile of the assessed population. If we interpret the difference between the average FPEs and scaled plausible values from the special studies as bias, then the results presented here suggest that the bias in imputing the mean of the assessed population is approximately 10 times as large as that in using the FPEs.

The composition of the special studies sample appears to include more able students than the average of the excluded student population. If this is true, then the difference between the (unobserved) ability of the entire excluded population and the FPEs (the bias in the FPEs) could be larger for the entire excluded population than for the special studies sample. Although the special studies provide no empirical evidence about the size of that bias, it is difficult to imagine that it could be larger than the bias implied by imputing the mean of the assessed population for these values.

These studies suggest that the calculator block and KaSA booklet accommodations, by themselves, will not change the number of included students enough to have a substantial impact on national parameter estimates. However, results for the FPE estimates on the entire excluded population do show nonnegligible impacts on national parameter estimates. This suggests that if accommodations to include more of the currently excluded students could be found, such accommodations could have a nonnegligible impact on national parameter estimates. Moreover, because FPEs appear to overestimate estimates based on scaled plausible values, the impact of including currently excluded students would likely be even larger than that estimated by the FPEs.

It is important to remember that these special studies are relatively small, and consequently their results have considerable sampling uncertainty that makes it difficult to draw sharp conclusions. The sampling uncertainty made it infeasible to carry out many analyses that would have been desirable. For example, it would be useful to see if patterns of bias were reasonably constant across states and across all reporting groups, but it was not meaningful to conduct these analyses. A fair conclusion is that the sampling uncertainty is so large that any conclusions drawn from this study must be done with extreme care.

It may be useful to question whether the concept of *full* population estimates is sensible. The reason is that the concept of full population estimates presupposes that there is (at least in theory) an assessment score for every student, including those who are currently excluded from the assessment. If there are students whom we could not conceive as participating in the assessment under any conditions, then the concept of "the assessment score they would have obtained if they had participated" may not make sense. Moreover, it is impossible that any empirical methods could be developed to impute assessment scores for a group that could never have assessment scores—no empirical information about assessment scores could exist for that group. Consequently, it will never be possible to validate methods of imputing assessment scores for a group that could never be assessed. One might therefore argue that a group that could never be assessed should be excluded from the definition of the population used to draw inferences. By redefining the population, efforts could focus on developing methods to include as many members of the (newly defined) population as possible in operational assessments and developing methods to impute scores for those excluded. Of course, there is a problem in identifying the group that should be defined as not (ever) assessable. Nevertheless, it may be worth attempting to develop at least provisional definitions of such a group.

This suggests a concept of *expanded* population estimates (rather than full population estimates) that corresponds to estimating the assessment scores that could be obtained by all students who could participate in the assessment under conditions of special accommodations. One virtue of this definition is that every student in the inference population could be assessed under *some* accommodations (including accommodations that might be infeasible under operational conditions because of time or cost constraints). Because it would be possible to obtain assessment scores for every student in the population, empirical methods could, in principle, be used to develop imputations for any students in this population who are excluded from the operational assessment (perhaps in special studies involving extensive accommodations). Moreover, it would be possible to empirically validate such methods.

Appendix A. Procedures for Calculating Full Population Estimates

McLaughlin (2000) introduced a method to estimate the achievement of the subset of the students with disabilities (SDs) and English language learners (ELLs) excluded by NAEP. The method relies on the NAEP SD and ELL questionnaires, descriptive surveys that are filled out by a teacher or knowledgeable staff person for each student with a disability and each English language learner selected to participate in NAEP—whether or not these students actually participate in NAEP or are excluded on the grounds that NAEP testing would be inappropriate for them.

The basic assumption is that excluded students in a given state with a particular profile based on student and school demographic characteristics and information from the SD and ELL questionnaires will, on average, be at the same achievement level as students with disabilities and English language learners in that state who participated in NAEP and had the same profile of demographic characteristics and information on the SD and ELL questionnaires. McLaughlin called this the profile matching method. Since the scores resulting from this procedure provide estimates that now include all of a state's SDs and ELLS, they are called *full population estimates* (FPEs).

No student takes the entire NAEP assessment. Instead, a student takes a random sample of blocks of items drawn from the entire item set for a given assessment. The items each student takes are used to compute five sets of what are called *plausible values*.[7] These are then used to compute estimates of performance for the entire population of students as well as congressionally mandated subgroups of students (e.g., males and females).

In computing the FPEs, plausible values for the composite NAEP scale in each grade and subject are computed first for all excluded ELLs in the NAEP public school sample, and second, separately, for all excluded SDs in the sample who are not also ELLs. Data for students who are neither ELL nor SD are not used in the process. The plausible values are constructed in three steps.

1. **Predictor preparation.** Predictive demographic information and questionnaire responses, which are available for both included and excluded ELLs and SDs, are extracted from the NAEP file and recoded to maximize predictive power. Stepwise regression is used to remove predictors possessing no significant power in predicting plausible values for included ELLs (or SDs) and to remove predictors that are too highly correlated with other predictors.
2. **Estimation of the mean expected score for each excluded student**. A single pooled within-state linear regression is carried out to estimate the coefficient for each of the predictors created in step 1 in predicting the scores of included ELLs

[7] The procedures in this paper used five plausible values, but the estimation procedure has been changed for the 2013 NAEP assessments and now generates 20 plausible values. Future versions of the software for generating FPEs will be updated to reflect this change.

(or SDs).[8] The regression intercept is adjusted separately for each state so that the mean predicted score for included ELLs (or SDs) matches their observed mean in each state. The resulting coefficients are used to impute an estimate for each excluded ELL (or SD).

3. **Estimation of imputation error variance and generation of five random plausible values for each excluded student**. Five plausible values are generated for each excluded student by adding to the estimate obtained in step 2 random normal deviates with three components of variance: (1) average variation among the five NAEP plausible values for included ELLs (or SDs), (2) average regression error due to the imperfect linear regression prediction in step 2, and (3) sampling error introduced in matching the included ELL (or SD) mean in the state.

One of the difficulties that the FPE procedure has had to deal with is that the set of questions that comprise the NAEP SD and ELL questionnaires have changed from year to year. As a result, the prediction equations change from NAEP administration to administration. While this fact does not diminish the utility of the FPE procedure, it does mean that the fit of regression results to the data can vary over time. Table A-1 below lists separately the variables used in the NAEP 2011 Grade 8 reading and mathematics FPE regressions for ELLs and SDs.

[8] A student's "score" is defined as the mean of the five plausible values for that student.

Table A-1. Variables Used in the Linear Regressions for Grade 8 Reading and Mathematics: 2011

		Mathematics Grade 8		Reading Grade 8	
Variable	**Description**	**SD**	**ELL**	**SD**	**ELL**
	Items from the ELL questionnaire				
XL04501	What is this student's ELL classification?				•
XL03801	How is student included in state assessment?		•		•
XL03901	Extended time (allowed for all subjects)				
XL03902	Small group (allowed for all subjects)				
XL03908	Test items read aloud in English				
XL03905	Breaks during testing (allowed for all subjects)				
XL03909	Must have an aide administer test		•		
XL03910	Cueing to stay on task				
XL03906	Bilingual dictionary w/out definitions in any language				
XL03911	Read directions aloud in Spanish				•
XL03912	Test items read aloud in Spanish (math & science)				•
XL03913	Spanish/English version of the test (math & science)		•		
XL03914	Student receives other accommodations				
XL04001	How should this student be included on NAEP test?				
XL04002	If student ineligible for NAEP, record admin. code				•
XL04101	How long has student been receiving instruction in English?				•
XL04201	Grade level of performance in NAEP subject		•		•
XL04301	Student's English proficiency: listening comprehension in English		•		•
XL04302	Student's English proficiency: Speaking English		•		
XL04303	Student's English proficiency: Reading English		•		•
XL04304	Student's English proficiency: Writing English		•		•
	Items from the SD questionnaire				
XS04701	Why is this student classified as SD?	•		•	
XS04801	How is student included in state assessment?	•	•	•	•
XS04901	Extended time (allowed for all subjects)		•	•	•
XS04902	Small group (allowed for all subjects)	•		•	
XS04907	Test items read aloud in English	•		•	•
XS04905	Breaks during testing (allowed for all subjects)				
XS04908	Must have an aide administer test		•		
XS04909	Responds orally to a scribe	•			
XS04910	Large-print version of the test	•			
XS04911	Magnification equipment	•			
XS04912	Uses a calculator for all sections (math only)	•			
XS04913	Uses template/special equip./preferential seating	•		•	

		Mathematics Grade 8		Reading Grade 8	
Variable	**Description**	**SD**	**ELL**	**SD**	**ELL**
XS04914	Cueing to stay on task				
XS04915	Presentation or response in braille				
XS04916	Presentation or response in sign language	•			
XS04917	Student receives other accommodations	•			
XS05001	How should this student be included on NAEP test?	•			
XS05002	If student ineligible for NAEP, record admin. code				
XS05101	Student's identified disability: Specific learning	•		•	•
XS05102	Student's identified disability: Hearing impairment	•			
XS05103	Student's identified disability: Visual impairment			•	
XS05105	Student's identified disability: Mental retardation	•	•	•	
XS05106	Student's identified disability: Emotional disturbance	•			
XS05107	Student's identified disability: Orthopedic impairment		•	•	
XS05108	Student's identified disability: Brain injury	•		•	
XS05109	Student's identified disability: Autism	•	•		
XS05110	Student's identified disability: Developmental delay				
XS05111	Student's identified disability: Other health	•	•		
XS05104	Student's identified disability: Speech impairment	•		•	•
XS05112	Student's identified disability: Other-write-in				
XS05201	Degree of student's disability	•	•	•	
XS05301	Grade level student performs in the NAEP subject	•	•	•	•
	Student and school characteristics				
IEP	Student classified as having a disability		•		•
DMIN	Student is not white	•		•	•
DSEX	Student gender	•	•	•	•
SLUNCH	National School Lunch Program eligibility	•	•	•	•
PCTBLK	School-level percentage of black students	•	•		•
PCTIND	School-level percentage of American Indian students	•	•		•
PCTHSP	School-level percentage of Hispanic students	•	•	•	
READVAR	School-level state test scores—Reading	•		•	•
MATHVAR	School-level state test scores—Math	•	•	•	
SENROL8	School enrollment		•		

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Reading and Mathematics Assessments.

Attachment A8 - NAEP 2013 Operations Report: Students Assessed, Accommodated, Excluded, and Absent, including total participation rates (available in both PDF and Excel formats)

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Total			831,930	792,560	631,303	97,349	91.9%	16,082	2.0%	6,419	41,407
Total	Grade 4	Mathematics	214,934	205,142	162,475	28,679	93.2%	3,116	1.5%	1,353	9,519
Total	Grade 4	Reading	216,426	206,683	164,835	25,572	92.1%	5,596	2.7%	1,375	9,305
Total	Grade 8	Mathematics	201,461	191,407	152,232	23,047	91.6%	2,912	1.5%	1,908	11,308
Total	Grade 8	Reading	199,109	189,328	151,761	20,051	90.7%	4,458	2.4%	1,783	11,275
Total TUDA			149,257	141,849	105,663	23,255	90.9%	4,139	2.9%	772	8,020
Total TUDA	Grade 4	Mathematics	38,864	37,044	27,356	6,969	92.7%	767	2.1%	238	1,714
Total TUDA	Grade 4	Reading	40,094	38,178	28,349	6,259	90.6%	1,571	4.1%	242	1,757
Total TUDA	Grade 8	Mathematics	34,758	32,942	24,684	5,124	90.5%	656	2.0%	153	2,325
Total TUDA	Grade 8	Reading	35,541	33,685	25,274	4,903	89.6%	1,145	3.4%	139	2,224
Total Public			818,369	779,146	619,153	96,739	91.9%	16,047	2.1%	6,349	40,858
Total Public	Grade 4	Mathematics	211,657	201,896	159,511	28,542	93.1%	3,114	1.5%	1,337	9,392
Total Public	Grade 4	Reading	213,046	203,340	161,770	25,454	92.1%	5,581	2.7%	1,361	9,174
Total Public	Grade 8	Mathematics	198,049	188,035	149,217	22,869	91.5%	2,904	1.5%	1,885	11,160
Total Public	Grade 8	Reading	195,617	185,875	148,655	19,874	90.7%	4,448	2.4%	1,766	11,132
Total Nonpublic			13,561	13,414	12,150	610	95.1%	35	0.3%	70	549
Total Nonpublic	Grade 4	Mathematics	3,277	3,246	2,964	137	95.5%	2	0.1%	16	127
Total Nonpublic	Grade 4	Reading	3,380	3,343	3,065	118	95.2%	15	0.4%	14	131
Total Nonpublic	Grade 8	Mathematics	3,412	3,372	3,015	178	94.7%	8	0.2%	23	148
Total Nonpublic	Grade 8	Reading	3,492	3,453	3,106	177	95.1%	10	0.3%	17	143
Alabama			12,683	11,964	10,630	578	93.7%	131	1.1%	92	533
Alabama	Grade 4	Mathematics	3,219	3,050	2,724	140	93.9%	32	1.0%	24	130
Alabama	Grade 4	Reading	3,351	3,175	2,828	169	94.4%	36	1.1%	23	119
Alabama	Grade 8	Mathematics	3,035	2,838	2,512	135	93.3%	29	1.0%	32	130
Alabama	Grade 8	Reading	3,078	2,901	2,566	134	93.1%	34	1.2%	13	154
Alaska			12,478	11,694	8,716	1,996	91.6%	150	1.3%	71	761
Alaska	Grade 4	Mathematics	3,131	2,946	2,173	535	91.9%	34	1.2%	15	189
Alaska	Grade 4	Reading	3,257	3,061	2,259	573	92.5%	44	1.4%	14	171
Alaska	Grade 8	Mathematics	3,002	2,821	2,095	470	90.9%	34	1.2%	22	200
Alaska	Grade 8	Reading	3,088	2,866	2,189	418	91.0%	38	1.3%	20	201
Arizona			13,394	12,559	10,359	1,358	93.3%	162	1.3%	95	585
Arizona	Grade 4	Mathematics	3,427	3,228	2,650	388	94.1%	38	1.2%	16	136
Arizona	Grade 4	Reading	3,531	3,338	2,738	414	94.4%	38	1.1%	25	123
Arizona	Grade 8	Mathematics	3,180	2,957	2,441	282	92.1%	41	1.4%	23	170
Arizona	Grade 8	Reading	3,256	3,036	2,530	274	92.4%	45	1.5%	31	156
Arkansas			13,327	12,475	9,903	1,717	93.1%	192	1.5%	34	629
Arkansas	Grade 4	Mathematics	3,423	3,223	2,554	464	93.6%	39	1.2%	9	157
Arkansas	Grade 4	Reading	3,560	3,348	2,646	507	94.2%	36	1.1%	6	153
Arkansas	Grade 8	Mathematics	3,161	2,939	2,350	388	93.2%	58	2.0%	6	137
Arkansas	Grade 8	Reading	3,183	2,965	2,353	358	91.4%	59	2.0%	13	182

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
California			35,116	33,587	28,238	2,752	92.3%	703	2.1%	173	1,721
California	Grade 4	Mathematics	9,000	8,598	7,292	739	93.4%	135	1.6%	32	400
California	Grade 4	Reading	9,258	8,856	7,473	721	92.5%	208	2.3%	37	417
California	Grade 8	Mathematics	8,353	7,981	6,648	681	91.8%	141	1.8%	58	453
California	Grade 8	Reading	8,505	8,152	6,825	611	91.2%	219	2.7%	46	451
Colorado			13,221	12,658	10,267	1,390	92.1%	157	1.2%	112	732
Colorado	Grade 4	Mathematics	3,400	3,276	2,626	365	91.3%	37	1.1%	26	222
Colorado	Grade 4	Reading	3,505	3,371	2,704	407	92.3%	51	1.5%	21	188
Colorado	Grade 8	Mathematics	3,126	2,968	2,424	317	92.4%	35	1.2%	29	163
Colorado	Grade 8	Reading	3,190	3,043	2,513	301	92.5%	34	1.1%	36	159
Connecticut			12,799	12,455	9,738	1,644	91.4%	222	1.8%	181	670
Connecticut	Grade 4	Mathematics	3,234	3,147	2,462	456	92.7%	44	1.4%	35	150
Connecticut	Grade 4	Reading	3,352	3,253	2,558	460	92.8%	52	1.6%	22	161
Connecticut	Grade 8	Mathematics	3,091	3,012	2,353	371	90.4%	64	2.1%	63	161
Connecticut	Grade 8	Reading	3,122	3,043	2,365	357	89.5%	62	2.0%	61	198
Delaware			13,362	12,992	10,177	1,537	90.2%	381	2.9%	53	844
Delaware	Grade 4	Mathematics	3,419	3,333	2,629	449	92.3%	71	2.1%	14	170
Delaware	Grade 4	Reading	3,524	3,432	2,761	330	90.1%	161	4.7%	9	171
Delaware	Grade 8	Mathematics	3,172	3,080	2,350	413	89.7%	39	1.3%	18	260
Delaware	Grade 8	Reading	3,247	3,147	2,437	345	88.4%	110	3.5%	12	243
District of Columbia			8,811	8,485	6,330	1,458	91.8%	123	1.4%	15	559
District of Columbia	Grade 4	Mathematics	2,279	2,209	1,698	373	93.8%	31	1.4%	5	102
District of Columbia	Grade 4	Reading	2,371	2,293	1,767	364	92.9%	37	1.6%	6	119
District of Columbia	Grade 8	Mathematics	2,065	1,986	1,430	365	90.4%	19	1.0%	3	169
District of Columbia	Grade 8	Reading	2,096	1,997	1,435	356	89.7%	36	1.8%	1	169
DoDEA/DDESS (co			12,708	11,435	9,514	978	91.8%	374	3.3%	36	533
DoDEA/DDESS (co	Grade 4	Mathematics	3,679	3,264	2,673	378	93.5%	54	1.7%	10	149
DoDEA/DDESS (co	Grade 4	Reading	3,777	3,383	2,806	233	89.8%	200	5.9%	13	131
DoDEA/DDESS (co	Grade 8	Mathematics	2,623	2,397	2,034	205	93.4%	27	1.1%	9	122
DoDEA/DDESS (co	Grade 8	Reading	2,629	2,391	2,001	162	90.5%	93	3.9%	4	131
Florida			26,966	25,547	18,534	4,827	91.4%	576	2.3%	286	1,324
Florida	Grade 4	Mathematics	6,919	6,556	4,621	1,476	93.0%	127	1.9%	67	265
Florida	Grade 4	Reading	7,082	6,739	4,759	1,424	91.7%	202	3.0%	59	295
Florida	Grade 8	Mathematics	6,423	6,073	4,543	948	90.4%	111	1.8%	91	380
Florida	Grade 8	Reading	6,542	6,179	4,611	979	90.5%	136	2.2%	69	384
Georgia			20,344	19,088	15,860	1,718	92.1%	418	2.2%	122	970
Georgia	Grade 4	Mathematics	5,252	4,960	4,134	499	93.4%	63	1.3%	34	230
Georgia	Grade 4	Reading	5,406	5,118	4,295	427	92.3%	178	3.5%	25	193
Georgia	Grade 8	Mathematics	4,802	4,464	3,634	456	91.6%	62	1.4%	33	279
Georgia	Grade 8	Reading	4,884	4,546	3,797	336	90.9%	115	2.5%	30	268

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Hawaii			13,546	12,894	10,398	1,358	91.2%	209	1.6%	74	855
Hawaii	Grade 4	Mathematics	3,482	3,301	2,762	326	93.5%	40	1.2%	20	153
Hawaii	Grade 4	Reading	3,593	3,401	2,816	323	92.3%	58	1.7%	16	188
Hawaii	Grade 8	Mathematics	3,212	3,078	2,391	353	89.1%	53	1.7%	22	259
Hawaii	Grade 8	Reading	3,259	3,114	2,429	356	89.4%	58	1.9%	16	255
Idaho			13,412	12,552	10,659	1,044	93.2%	176	1.4%	72	601
Idaho	Grade 4	Mathematics	3,479	3,269	2,780	291	93.9%	43	1.3%	13	142
Idaho	Grade 4	Reading	3,576	3,377	2,839	317	93.5%	54	1.6%	14	153
Idaho	Grade 8	Mathematics	3,147	2,922	2,509	216	93.3%	29	1.0%	19	149
Idaho	Grade 8	Reading	3,210	2,984	2,531	220	92.2%	50	1.7%	26	157
Illinois			20,005	19,387	15,302	2,812	93.4%	241	1.2%	138	894
Illinois	Grade 4	Mathematics	5,081	4,910	3,791	806	93.6%	54	1.1%	31	228
Illinois	Grade 4	Reading	5,248	5,080	4,018	748	93.8%	62	1.2%	34	218
Illinois	Grade 8	Mathematics	4,780	4,636	3,686	650	93.5%	54	1.2%	39	207
Illinois	Grade 8	Reading	4,896	4,761	3,807	608	92.7%	71	1.5%	34	241
Indiana			12,849	12,243	9,560	1,715	92.1%	228	1.9%	118	622
Indiana	Grade 4	Mathematics	3,340	3,202	2,496	508	93.8%	44	1.4%	11	143
Indiana	Grade 4	Reading	3,456	3,306	2,547	493	92.0%	84	2.5%	28	154
Indiana	Grade 8	Mathematics	2,994	2,830	2,251	327	91.1%	43	1.5%	46	163
Indiana	Grade 8	Reading	3,059	2,905	2,266	387	91.3%	57	2.0%	33	162
Iowa			12,464	11,998	9,694	1,516	93.4%	119	1.0%	84	585
Iowa	Grade 4	Mathematics	3,092	2,989	2,414	408	94.4%	22	0.7%	14	131
Iowa	Grade 4	Reading	3,192	3,092	2,485	423	94.0%	35	1.1%	11	138
Iowa	Grade 8	Mathematics	3,064	2,942	2,382	355	93.0%	23	0.8%	29	153
Iowa	Grade 8	Reading	3,116	2,975	2,413	330	92.2%	39	1.3%	30	163
Kansas			13,582	12,971	10,437	1,588	92.7%	220	1.7%	73	653
Kansas	Grade 4	Mathematics	3,439	3,283	2,628	438	93.4%	48	1.5%	8	161
Kansas	Grade 4	Reading	3,541	3,395	2,711	453	93.2%	64	1.9%	20	147
Kansas	Grade 8	Mathematics	3,269	3,121	2,552	334	92.5%	51	1.6%	28	156
Kansas	Grade 8	Reading	3,333	3,172	2,546	363	91.7%	57	1.8%	17	189
Kentucky			18,070	17,228	14,230	1,615	92.0%	462	2.7%	68	853
Kentucky	Grade 4	Mathematics	4,653	4,450	3,672	479	93.3%	68	1.5%	17	214
Kentucky	Grade 4	Reading	4,798	4,584	3,819	383	91.7%	159	3.5%	24	199
Kentucky	Grade 8	Mathematics	4,285	4,076	3,356	405	92.3%	77	1.9%	16	222
Kentucky	Grade 8	Reading	4,334	4,118	3,383	348	90.6%	158	3.8%	11	218
Louisiana			13,274	12,189	9,515	1,843	93.2%	141	1.2%	95	595
Louisiana	Grade 4	Mathematics	3,331	3,117	2,415	500	93.5%	31	1.0%	31	140
Louisiana	Grade 4	Reading	3,442	3,192	2,420	567	93.6%	39	1.2%	31	135
Louisiana	Grade 8	Mathematics	3,237	2,924	2,311	406	92.9%	34	1.2%	19	154
Louisiana	Grade 8	Reading	3,264	2,956	2,369	370	92.7%	37	1.3%	14	166

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Maine			12,742	12,276	9,413	1,821	91.5%	206	1.7%	141	695
Maine	Grade 4	Mathematics	3,376	3,279	2,502	505	91.7%	69	2.1%	31	172
Maine	Grade 4	Reading	3,474	3,370	2,548	551	92.0%	58	1.7%	43	170
Maine	Grade 8	Mathematics	2,918	2,790	2,164	384	91.3%	37	1.3%	34	171
Maine	Grade 8	Reading	2,974	2,837	2,199	381	90.9%	42	1.5%	33	182
Maryland			18,436	17,699	13,566	1,749	86.5%	1,247	7.0%	130	1,007
Maryland	Grade 4	Mathematics	4,748	4,545	3,461	773	93.2%	52	1.1%	33	226
Maryland	Grade 4	Reading	4,900	4,702	3,597	231	81.4%	640	13.6%	31	203
Maryland	Grade 8	Mathematics	4,354	4,188	3,237	523	89.8%	68	1.6%	39	321
Maryland	Grade 8	Reading	4,434	4,264	3,271	222	81.9%	487	11.4%	27	257
Massachusetts			20,140	19,612	14,747	2,996	90.5%	524	2.7%	306	1,039
Massachusetts	Grade 4	Mathematics	5,155	5,012	3,710	862	91.2%	134	2.7%	61	245
Massachusetts	Grade 4	Reading	5,318	5,171	3,978	721	90.9%	165	3.2%	80	227
Massachusetts	Grade 8	Mathematics	4,786	4,665	3,462	730	89.9%	102	2.2%	83	288
Massachusetts	Grade 8	Reading	4,881	4,764	3,597	683	89.8%	123	2.6%	82	279
Michigan			17,938	16,818	13,377	1,703	89.7%	600	3.6%	98	1,040
Michigan	Grade 4	Mathematics	4,620	4,340	3,455	474	90.5%	124	2.9%	16	271
Michigan	Grade 4	Reading	4,753	4,450	3,573	412	89.6%	198	4.4%	20	247
Michigan	Grade 8	Mathematics	4,223	3,952	3,133	414	89.8%	103	2.6%	36	266
Michigan	Grade 8	Reading	4,342	4,076	3,216	403	88.8%	175	4.3%	26	256
Minnesota			13,038	12,452	10,297	1,091	91.5%	252	2.0%	171	641
Minnesota	Grade 4	Mathematics	3,491	3,324	2,780	333	93.7%	45	1.4%	27	139
Minnesota	Grade 4	Reading	3,609	3,455	2,867	323	92.3%	96	2.8%	16	153
Minnesota	Grade 8	Mathematics	2,939	2,806	2,290	232	89.9%	49	1.7%	62	173
Minnesota	Grade 8	Reading	2,999	2,867	2,360	203	89.4%	62	2.2%	66	176
Mississippi			13,183	12,512	10,980	768	93.9%	84	0.7%	92	588
Mississippi	Grade 4	Mathematics	3,336	3,193	2,827	198	94.7%	24	0.8%	16	128
Mississippi	Grade 4	Reading	3,447	3,297	2,905	210	94.5%	17	0.5%	19	146
Mississippi	Grade 8	Mathematics	3,177	2,982	2,596	180	93.1%	23	0.8%	30	153
Mississippi	Grade 8	Reading	3,223	3,040	2,652	180	93.2%	20	0.7%	27	161
Missouri			13,498	12,463	10,382	1,247	93.3%	152	1.2%	83	599
Missouri	Grade 4	Mathematics	3,612	3,294	2,775	329	94.2%	43	1.3%	9	138
Missouri	Grade 4	Reading	3,712	3,401	2,833	367	94.1%	44	1.3%	27	130
Missouri	Grade 8	Mathematics	3,063	2,860	2,377	283	93.0%	36	1.3%	19	145
Missouri	Grade 8	Reading	3,111	2,908	2,397	268	91.6%	29	1.0%	28	186
Montana			13,202	12,666	10,608	903	90.9%	274	2.2%	78	803
Montana	Grade 4	Mathematics	3,350	3,216	2,709	246	91.9%	58	1.8%	10	193
Montana	Grade 4	Reading	3,470	3,354	2,844	222	91.4%	100	3.0%	13	175
Montana	Grade 8	Mathematics	3,170	3,028	2,515	234	90.8%	45	1.5%	23	211
Montana	Grade 8	Reading	3,212	3,068	2,540	201	89.3%	71	2.3%	32	224

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Nebraska			13,283	12,763	10,239	1,492	91.9%	326	2.6%	54	652
Nebraska	Grade 4	Mathematics	3,468	3,327	2,680	442	93.8%	56	1.7%	10	139
Nebraska	Grade 4	Reading	3,564	3,426	2,712	452	92.4%	123	3.6%	10	129
Nebraska	Grade 8	Mathematics	3,091	2,965	2,395	323	91.7%	55	1.9%	17	175
Nebraska	Grade 8	Reading	3,160	3,045	2,452	275	89.6%	92	3.0%	17	209
Nevada			13,950	13,042	9,987	2,127	92.9%	158	1.2%	93	677
Nevada	Grade 4	Mathematics	3,541	3,306	2,376	745	94.4%	47	1.4%	13	125
Nevada	Grade 4	Reading	3,654	3,413	2,477	726	93.8%	46	1.3%	13	151
Nevada	Grade 8	Mathematics	3,336	3,127	2,555	319	91.9%	31	1.0%	27	195
Nevada	Grade 8	Reading	3,419	3,196	2,579	337	91.2%	34	1.1%	40	206
New Hampshire			13,223	12,906	10,033	1,704	90.9%	245	1.9%	240	684
New Hampshire	Grade 4	Mathematics	3,357	3,283	2,585	456	92.6%	40	1.2%	50	152
New Hampshire	Grade 4	Reading	3,460	3,380	2,681	405	91.3%	82	2.4%	53	159
New Hampshire	Grade 8	Mathematics	3,177	3,090	2,345	459	90.7%	33	1.1%	71	182
New Hampshire	Grade 8	Reading	3,229	3,153	2,422	384	89.0%	90	2.9%	66	191
New Jersey			13,124	12,809	9,936	1,828	91.8%	221	1.7%	256	568
New Jersey	Grade 4	Mathematics	3,328	3,249	2,554	490	93.7%	39	1.2%	33	133
New Jersey	Grade 4	Reading	3,449	3,361	2,668	466	93.2%	55	1.6%	38	134
New Jersey	Grade 8	Mathematics	3,147	3,070	2,292	488	90.6%	52	1.7%	83	155
New Jersey	Grade 8	Reading	3,200	3,129	2,422	384	89.7%	75	2.4%	102	146
New Mexico			16,480	15,406	12,241	2,025	92.6%	213	1.4%	87	840
New Mexico	Grade 4	Mathematics	4,172	3,908	2,990	673	93.7%	51	1.3%	12	182
New Mexico	Grade 4	Reading	4,300	4,049	3,200	578	93.3%	40	1.0%	17	214
New Mexico	Grade 8	Mathematics	3,965	3,695	2,942	435	91.4%	57	1.5%	28	233
New Mexico	Grade 8	Reading	4,043	3,754	3,109	339	91.8%	65	1.7%	30	211
New York			17,873	17,470	12,426	3,391	90.5%	238	1.4%	433	982
New York	Grade 4	Mathematics	4,509	4,423	3,138	891	91.1%	56	1.3%	128	210
New York	Grade 4	Reading	4,644	4,546	3,277	888	91.6%	63	1.4%	119	199
New York	Grade 8	Mathematics	4,336	4,233	3,034	766	89.8%	70	1.7%	86	277
New York	Grade 8	Reading	4,384	4,268	2,977	846	89.6%	49	1.1%	100	296
North Carolina			18,874	18,076	14,528	2,133	92.2%	261	1.4%	210	944
North Carolina	Grade 4	Mathematics	4,839	4,621	3,688	609	93.0%	62	1.3%	54	208
North Carolina	Grade 4	Reading	4,982	4,787	3,897	567	93.3%	73	1.5%	50	200
North Carolina	Grade 8	Mathematics	4,480	4,296	3,420	502	91.3%	56	1.3%	53	265
North Carolina	Grade 8	Reading	4,573	4,372	3,523	455	91.0%	70	1.6%	53	271
North Dakota			15,046	14,203	11,759	1,300	91.9%	490	3.4%	42	612
North Dakota	Grade 4	Mathematics	3,700	3,502	2,942	319	93.1%	90	2.6%	17	134
North Dakota	Grade 4	Reading	3,845	3,637	3,055	306	92.4%	146	4.0%	7	123
North Dakota	Grade 8	Mathematics	3,716	3,492	2,828	386	92.0%	108	3.1%	7	163
North Dakota	Grade 8	Reading	3,785	3,572	2,934	289	90.2%	146	4.1%	11	192

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Ohio			18,579	17,470	13,246	2,684	91.2%	439	2.5%	110	991
Ohio	Grade 4	Mathematics	4,682	4,409	3,403	653	92.0%	98	2.2%	17	238
Ohio	Grade 4	Reading	4,809	4,552	3,493	667	91.4%	150	3.3%	20	222
Ohio	Grade 8	Mathematics	4,491	4,205	3,157	671	91.0%	77	1.8%	37	263
Ohio	Grade 8	Reading	4,597	4,304	3,193	693	90.3%	114	2.6%	36	268
Oklahoma			13,603	12,713	10,141	1,610	92.4%	208	1.6%	98	656
Oklahoma	Grade 4	Mathematics	3,595	3,384	2,688	448	92.7%	62	1.8%	14	172
Oklahoma	Grade 4	Reading	3,689	3,464	2,769	455	93.1%	59	1.7%	16	165
Oklahoma	Grade 8	Mathematics	3,126	2,903	2,274	382	91.5%	48	1.7%	35	164
Oklahoma	Grade 8	Reading	3,193	2,962	2,410	325	92.3%	39	1.3%	33	155
Oregon			13,495	12,780	10,244	1,465	91.6%	252	2.0%	138	681
Oregon	Grade 4	Mathematics	3,549	3,372	2,605	496	92.0%	80	2.4%	25	166
Oregon	Grade 4	Reading	3,665	3,505	2,832	384	91.8%	84	2.4%	23	182
Oregon	Grade 8	Mathematics	3,123	2,915	2,383	283	91.5%	45	1.5%	41	163
Oregon	Grade 8	Reading	3,158	2,988	2,424	302	91.2%	43	1.4%	49	170
Pennsylvania			17,781	17,098	13,269	2,331	91.2%	398	2.3%	152	948
Pennsylvania	Grade 4	Mathematics	4,531	4,375	3,480	566	92.5%	93	2.1%	23	213
Pennsylvania	Grade 4	Reading	4,648	4,478	3,565	560	92.1%	121	2.7%	21	211
Pennsylvania	Grade 8	Mathematics	4,260	4,077	3,073	612	90.4%	86	2.1%	46	260
Pennsylvania	Grade 8	Reading	4,342	4,168	3,151	593	89.8%	98	2.4%	62	264
Puerto Rico			10,968	10,530	7,407	2,419	93.3%	13	0.1%	130	561
Puerto Rico	Grade 4	Mathematics	5,086	4,906	3,381	1,239	94.2%	11	0.2%	55	220
Puerto Rico	Grade 8	Mathematics	5,882	5,624	4,026	1,180	92.6%	2	0.0%	75	341
Rhode Island			13,410	13,003	10,327	1,775	93.1%	160	1.2%	54	687
Rhode Island	Grade 4	Mathematics	3,393	3,291	2,617	475	94.0%	38	1.2%	10	151
Rhode Island	Grade 4	Reading	3,483	3,368	2,747	409	93.7%	43	1.3%	8	161
Rhode Island	Grade 8	Mathematics	3,242	3,155	2,494	437	92.9%	36	1.1%	15	173
Rhode Island	Grade 8	Reading	3,292	3,189	2,469	454	91.7%	43	1.3%	21	202
South Carolina			12,903	12,158	10,229	1,121	93.4%	183	1.5%	74	551
South Carolina	Grade 4	Mathematics	3,214	3,053	2,575	327	95.1%	31	1.0%	12	108
South Carolina	Grade 4	Reading	3,320	3,150	2,618	309	92.9%	58	1.8%	17	148
South Carolina	Grade 8	Mathematics	3,159	2,954	2,478	274	93.2%	37	1.3%	24	141
South Carolina	Grade 8	Reading	3,210	3,001	2,558	211	92.3%	57	1.9%	21	154
South Dakota			13,430	12,862	10,942	1,045	93.2%	253	2.0%	20	602
South Dakota	Grade 4	Mathematics	3,438	3,301	2,779	321	93.9%	50	1.5%	2	149
South Dakota	Grade 4	Reading	3,531	3,388	2,867	295	93.3%	78	2.3%	7	141
South Dakota	Grade 8	Mathematics	3,194	3,040	2,569	260	93.1%	42	1.4%	7	162
South Dakota	Grade 8	Reading	3,267	3,133	2,727	169	92.4%	83	2.6%	4	150
Tennessee			13,280	12,539	10,266	1,240	91.8%	299	2.4%	115	619
Tennessee	Grade 4	Mathematics	3,371	3,193	2,592	378	93.0%	42	1.3%	31	150

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Tennessee	Grade 4	Reading	3,504	3,305	2,654	396	92.3%	105	3.2%	27	123
Tennessee	Grade 8	Mathematics	3,194	3,012	2,485	258	91.1%	56	1.9%	34	179
Tennessee	Grade 8	Reading	3,211	3,029	2,535	208	90.6%	96	3.2%	23	167
Texas			36,372	34,425	25,227	6,055	90.9%	1,275	3.7%	245	1,623
Texas	Grade 4	Mathematics	9,201	8,765	6,084	2,132	93.7%	169	1.9%	72	308
Texas	Grade 4	Reading	9,513	9,043	6,227	1,803	88.8%	644	7.1%	60	309
Texas	Grade 8	Mathematics	8,756	8,223	6,365	1,138	91.2%	169	2.1%	58	493
Texas	Grade 8	Reading	8,902	8,394	6,551	982	89.7%	293	3.5%	55	513
Utah			13,977	13,353	10,914	1,284	91.4%	301	2.3%	125	729
Utah	Grade 4	Mathematics	3,551	3,376	2,753	406	93.6%	43	1.3%	17	157
Utah	Grade 4	Reading	3,655	3,484	2,839	322	90.7%	110	3.2%	26	187
Utah	Grade 8	Mathematics	3,347	3,198	2,612	295	90.9%	47	1.5%	45	199
Utah	Grade 8	Reading	3,424	3,295	2,710	261	90.2%	101	3.1%	37	186
Vermont			12,209	11,842	9,372	1,666	93.2%	128	1.1%	64	612
Vermont	Grade 4	Mathematics	3,000	2,903	2,300	421	93.7%	40	1.4%	14	128
Vermont	Grade 4	Reading	3,094	3,007	2,397	428	93.9%	35	1.2%	19	128
Vermont	Grade 8	Mathematics	3,025	2,933	2,309	422	93.1%	25	0.9%	10	167
Vermont	Grade 8	Reading	3,090	2,999	2,366	395	92.1%	28	0.9%	21	189
Virginia			13,289	12,696	10,357	1,409	92.7%	175	1.4%	128	627
Virginia	Grade 4	Mathematics	3,330	3,204	2,575	403	92.9%	49	1.5%	22	155
Virginia	Grade 4	Reading	3,448	3,304	2,717	375	93.6%	50	1.5%	21	141
Virginia	Grade 8	Mathematics	3,227	3,066	2,493	340	92.4%	33	1.1%	43	157
Virginia	Grade 8	Reading	3,284	3,122	2,572	291	91.7%	43	1.4%	42	174
Washington			13,581	13,085	10,401	1,449	90.6%	311	2.4%	161	763
Washington	Grade 4	Mathematics	3,554	3,424	2,696	454	92.0%	75	2.2%	36	163
Washington	Grade 4	Reading	3,694	3,561	2,850	413	91.6%	99	2.8%	39	160
Washington	Grade 8	Mathematics	3,139	3,016	2,380	313	89.3%	62	2.1%	51	210
Washington	Grade 8	Reading	3,194	3,084	2,475	269	89.0%	75	2.4%	35	230
West Virginia			12,831	12,162	10,197	974	91.9%	214	1.8%	84	693
West Virginia	Grade 4	Mathematics	3,156	3,007	2,517	280	93.0%	54	1.8%	11	145
West Virginia	Grade 4	Reading	3,264	3,119	2,611	256	91.9%	55	1.8%	26	171
West Virginia	Grade 8	Mathematics	3,178	2,987	2,493	230	91.2%	48	1.6%	20	196
West Virginia	Grade 8	Reading	3,233	3,049	2,576	208	91.3%	57	1.9%	27	181
Wisconsin			17,764	17,167	12,930	2,893	92.2%	391	2.3%	92	861
Wisconsin	Grade 4	Mathematics	4,445	4,297	3,246	755	93.1%	98	2.3%	14	184
Wisconsin	Grade 4	Reading	4,540	4,407	3,287	788	92.5%	102	2.3%	21	209
Wisconsin	Grade 8	Mathematics	4,346	4,203	3,174	669	91.4%	91	2.2%	28	241
Wisconsin	Grade 8	Reading	4,433	4,260	3,223	681	91.6%	100	2.3%	29	227
Wyoming			13,907	13,191	10,718	1,527	92.8%	161	1.2%	52	733
Wyoming	Grade 4	Mathematics	3,525	3,358	2,740	406	93.7%	34	1.0%	10	168

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	**Actual Student Sample**	Standard	Accommodated	**Total Participation Rate**	Excluded	%Excluded	Parent Refusal	Other Absence
Wyoming	Grade 4	Reading	3,624	3,462	2,817	410	93.2%	43	1.2%	16	176
Wyoming	Grade 8	Mathematics	3,348	3,138	2,543	352	92.3%	47	1.5%	13	183
Wyoming	Grade 8	Reading	3,410	3,233	2,618	359	92.1%	37	1.1%	13	206

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

					Assessed			Not Assessed			
			Original Student Sample	Actual Student Sample	Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Total			831,930	792,560	631,303	97,349	91.9%	16,082	2.0%	6,419	41,407
Total	Grade 4	Mathematics	214,934	205,142	162,475	28,679	93.2%	3,116	1.5%	1,353	9,519
Total	Grade 4	Reading	216,426	206,683	164,835	25,572	92.1%	5,596	2.7%	1,375	9,305
Total	Grade 8	Mathematics	201,461	191,407	152,232	23,047	91.6%	2,912	1.5%	1,908	11,308
Total	Grade 8	Reading	199,109	189,328	151,761	20,051	90.7%	4,458	2.4%	1,783	11,275
Total Public			818,369	779,146	619,153	96,739	91.9%	16,047	2.1%	6,349	40,858
Total Public	Grade 4	Mathematics	211,657	201,896	159,511	28,542	93.1%	3,114	1.5%	1,337	9,392
Total Public	Grade 4	Reading	213,046	203,340	161,770	25,454	92.1%	5,581	2.7%	1,361	9,174
Total Public	Grade 8	Mathematics	198,049	188,035	149,217	22,869	91.5%	2,904	1.5%	1,885	11,160
Total Public	Grade 8	Reading	195,617	185,875	148,655	19,874	90.7%	4,448	2.4%	1,766	11,132
Total Nonpublic			13,561	13,414	12,150	610	95.1%	35	0.3%	70	549
Total Nonpublic	Grade 4	Mathematics	3,277	3,246	2,964	137	95.5%	2	0.1%	16	127
Total Nonpublic	Grade 4	Reading	3,380	3,343	3,065	118	95.2%	15	0.4%	14	131
Total Nonpublic	Grade 8	Mathematics	3,412	3,372	3,015	178	94.7%	8	0.2%	23	148
Total Nonpublic	Grade 8	Reading	3,492	3,453	3,106	177	95.1%	10	0.3%	17	143

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent - by Race/Ethnicity

			Original Student Sample	Actual Student Sample	Assessed			Not Assessed			
					Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Total			831,930	792,560	631,303	97,349	91.9%	16,082	2.0%	6,419	41,407
Total	Grade 4	Mathematics	214,934	205,142	162,475	28,679	93.2%	3,116	1.5%	1,353	9,519
Total	Grade 4	Reading	216,426	206,683	164,835	25,572	92.1%	5,596	2.7%	1,375	9,305
Total	Grade 8	Mathematics	201,461	191,407	152,232	23,047	91.6%	2,912	1.5%	1,908	11,308
Total	Grade 8	Reading	199,109	189,328	151,761	20,051	90.7%	4,458	2.4%	1,783	11,275
White, not Hispanic			436,625	419,910	348,706	38,446	92.2%	6,579	1.6%	4,556	21,623
White, not Hispanic	Grade 4	Mathematics	110,037	105,961	88,535	10,221	93.2%	1,333	1.3%	892	4,980
White, not Hispanic	Grade 4	Reading	113,628	109,518	91,538	9,977	92.7%	2,117	1.9%	927	4,959
White, not Hispanic	Grade 8	Mathematics	105,206	100,919	83,094	9,422	91.7%	1,322	1.3%	1,369	5,712
White, not Hispanic	Grade 8	Reading	107,754	103,512	85,539	8,826	91.2%	1,807	1.7%	1,368	5,972
Black or African American, not Hispanic			146,123	135,961	107,176	16,767	91.2%	3,681	2.7%	571	7,766
Black or African American, not Hispanic	Grade 4	Mathematics	37,044	34,644	27,698	4,407	92.7%	684	2.0%	151	1,704
Black or African American, not Hispanic	Grade 4	Reading	38,003	35,534	28,526	3,976	91.5%	1,254	3.5%	148	1,630
Black or African American, not Hispanic	Grade 8	Mathematics	35,452	32,824	25,376	4,434	90.8%	614	1.9%	142	2,258
Black or African American, not Hispanic	Grade 8	Reading	35,624	32,959	25,576	3,950	89.6%	1,129	3.4%	130	2,174
Hispanic, of any race			169,977	161,673	115,456	33,024	91.8%	4,147	2.6%	832	8,214
Hispanic, of any race	Grade 4	Mathematics	46,908	44,698	30,362	11,403	93.4%	739	1.7%	220	1,974
Hispanic, of any race	Grade 4	Reading	43,265	41,225	28,443	9,136	91.2%	1,691	4.1%	199	1,756
Hispanic, of any race	Grade 8	Mathematics	42,557	40,404	30,038	7,080	91.9%	665	1.6%	252	2,369
Hispanic, of any race	Grade 8	Reading	37,247	35,346	26,613	5,405	90.6%	1,052	3.0%	161	2,115
Asian, not Hispanic			33,618	32,702	26,743	3,774	93.3%	869	2.7%	214	1,102
Asian, not Hispanic	Grade 4	Mathematics	8,891	8,615	6,981	1,125	94.1%	185	2.1%	41	283
Asian, not Hispanic	Grade 4	Reading	9,169	8,883	7,207	1,096	93.5%	252	2.8%	44	284
Asian, not Hispanic	Grade 8	Mathematics	7,786	7,594	6,265	818	93.3%	178	2.3%	65	268
Asian, not Hispanic	Grade 8	Reading	7,772	7,610	6,290	735	92.3%	254	3.3%	64	267
American Indian or Alaska Native, not Hispanic			17,483	16,046	12,037	2,457	90.3%	357	2.2%	75	1,120
American Indian or Alaska Native, not Hispanic	Grade 4	Mathematics	4,359	4,030	3,008	698	92.0%	72	1.8%	15	237
American Indian or Alaska Native, not Hispanic	Grade 4	Reading	4,556	4,235	3,215	610	90.3%	127	3.0%	16	267
American Indian or Alaska Native, not Hispanic	Grade 8	Mathematics	4,218	3,839	2,816	635	89.9%	63	1.6%	24	301
American Indian or Alaska Native, not Hispanic	Grade 8	Reading	4,350	3,942	2,998	514	89.1%	95	2.4%	20	315
Native Hawaiian or Pacific Islander, not Hispanic			6,996	6,602	5,018	921	90.0%	120	1.8%	27	516
Native Hawaiian or Pacific Islander, not Hispanic	Grade 4	Mathematics	1,803	1,697	1,329	248	92.9%	27	1.6%	5	88
Native Hawaiian or Pacific Islander, not Hispanic	Grade 4	Reading	1,794	1,675	1,306	214	90.7%	31	1.9%	6	118
Native Hawaiian or Pacific Islander, not Hispanic	Grade 8	Mathematics	1,679	1,603	1,178	241	88.5%	27	1.7%	10	147
Native Hawaiian or Pacific Islander, not Hispanic	Grade 8	Reading	1,720	1,627	1,205	218	87.5%	35	2.2%	6	163
Two or more races, not Hispanic			20,850	19,666	16,167	1,960	92.2%	329	1.7%	144	1,066
Two or more races, not Hispanic	Grade 4	Mathematics	5,818	5,497	4,562	577	93.5%	76	1.4%	29	253
Two or more races, not Hispanic	Grade 4	Reading	5,932	5,613	4,600	563	92.0%	124	2.2%	35	291
Two or more races, not Hispanic	Grade 8	Mathematics	4,509	4,224	3,465	417	91.9%	43	1.0%	46	253
Two or more races, not Hispanic	Grade 8	Reading	4,591	4,332	3,540	403	91.0%	86	2.0%	34	269
Information unavailable			258	0	0	0	#DIV/0!	0	#DIV/0!	0	0
Information unavailable	Grade 4	Mathematics	74	0	0	0	#DIV/0!	0	#DIV/0!	0	0
Information unavailable	Grade 4	Reading	79	0	0	0	#DIV/0!	0	#DIV/0!	0	0
Information unavailable	Grade 8	Mathematics	54	0	0	0	#DIV/0!	0	#DIV/0!	0	0
Information unavailable	Grade 8	Reading	51	0	0	0	#DIV/0!	0	#DIV/0!	0	0

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent - by NSLP

			Original Student Sample	Actual Student Sample	Assessed			Not Assessed			
					Standard	Accommodated	Total Participation Rate	Excluded	%Excluded	Parent Refusal	Other Absence
Total			831,930	792,560	631,303	97,349	91.9%	16,082	2.0%	6,419	41,407
Total	Grade 4	Mathematics	214,934	205,142	162,475	28,679	93.2%	3,116	1.5%	1,353	9,519
Total	Grade 4	Reading	216,426	206,683	164,835	25,572	92.1%	5,596	2.7%	1,375	9,305
Total	Grade 8	Mathematics	201,461	191,407	152,232	23,047	91.6%	2,912	1.5%	1,908	11,308
Total	Grade 8	Reading	199,109	189,328	151,761	20,051	90.7%	4,458	2.4%	1,783	11,275
Student not eligible			356,447	346,886	296,699	26,019	93.0%	4,490	1.3%	4,184	15,494
Student not eligible	Grade 4	Mathematics	87,476	85,215	73,175	6,880	93.9%	853	1.0%	811	3,496
Student not eligible	Grade 4	Reading	90,213	87,979	75,728	6,570	93.5%	1,395	1.6%	810	3,476
Student not eligible	Grade 8	Mathematics	88,491	85,949	73,095	6,461	92.6%	922	1.1%	1,282	4,189
Student not eligible	Grade 8	Reading	90,267	87,743	74,701	6,108	92.1%	1,320	1.5%	1,281	4,333
Free lunch			411,397	385,372	284,841	65,218	90.8%	10,402	2.7%	1,826	23,085
Free lunch	Grade 4	Mathematics	111,232	104,704	76,740	20,106	92.5%	2,042	2.0%	457	5,359
Free lunch	Grade 4	Reading	109,428	102,921	76,065	17,482	90.9%	3,732	3.6%	474	5,168
Free lunch	Grade 8	Mathematics	97,514	90,889	67,142	15,055	90.4%	1,810	2.0%	497	6,385
Free lunch	Grade 8	Reading	93,223	86,858	64,894	12,575	89.2%	2,818	3.2%	398	6,173
Reduced price lunch			38,723	37,517	30,304	4,500	92.8%	707	1.9%	248	1,758
Reduced price lunch	Grade 4	Mathematics	9,577	9,307	7,503	1,205	93.6%	145	1.6%	49	405
Reduced price lunch	Grade 4	Reading	9,968	9,682	7,805	1,177	92.8%	234	2.4%	56	410
Reduced price lunch	Grade 8	Mathematics	9,561	9,242	7,442	1,130	92.8%	131	1.4%	75	464
Reduced price lunch	Grade 8	Reading	9,617	9,286	7,554	988	92.0%	197	2.1%	68	479
School not participating			21,708	20,260	17,449	1,387	93.0%	398	2.0%	130	896
School not participating	Grade 4	Mathematics	5,855	5,400	4,642	448	94.3%	54	1.0%	29	227
School not participating	Grade 4	Reading	6,007	5,564	4,802	298	91.7%	213	3.8%	30	221
School not participating	Grade 8	Mathematics	4,889	4,617	3,985	331	93.5%	34	0.7%	42	225
School not participating	Grade 8	Reading	4,957	4,679	4,020	310	92.5%	97	2.1%	29	223
School refused			165	163	150	8	96.9%	0	0.0%	1	4
School refused	Grade 4	Mathematics	60	58	52	4	96.6%	0	0.0%	1	1
School refused	Grade 4	Reading	64	64	59	3	96.9%	0	0.0%	0	2
School refused	Grade 8	Mathematics	20	20	20	0	100.0%	0	0.0%	0	0
School refused	Grade 8	Reading	21	21	19	1	95.2%	0	0.0%	0	1
Information unavailable			3,490	2,362	1,860	217	87.9%	85	3.6%	30	170
Information unavailable	Grade 4	Mathematics	734	458	363	36	87.1%	22	4.8%	6	31
Information unavailable	Grade 4	Reading	746	473	376	42	88.4%	22	4.7%	5	28
Information unavailable	Grade 8	Mathematics	986	690	548	70	89.6%	15	2.2%	12	45
Information unavailable	Grade 8	Reading	1,024	741	573	69	86.6%	26	3.5%	7	66

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.

Average scale scores and percentages for reading, grade 8 by status as English Language Learner, 3 categories [ELL3], year and jurisdiction: 2013, 2011, 2009, 2007, 2005, and 2003

		Yes		No		Formerly ELL	
Year	**Jurisdiction**	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage
2013	National public	225	5	269	92	257	3
	California	220	12	269	74	256	14
	Florida	226	4	268	88	262	7
	New York	215	6	270	94	‡	#
	Texas	227	7	267	89	255	4
2011	National public	223	5	266	91	254	4
	California	220	17	264	69	252	15
	Florida	225	4	264	88	261	8
	New York	216	5	268	95	‡	#
	Texas	225	7	265	89	247	3
2009	National public	219	5	265	92	253	3
	California	215	19	263	68	252	13
	Florida	233	3	266	94	257	3
	New York	213	3	266	97	‡	#
	Texas	216	6	263	91	255	3
2007	National public	222	6	263	93	254	2
	California	219	21	261	71	255	8
	Florida	232	3	261	95	250	2
	New York	211	3	265	97	‡	#
	Texas	215	6	264	91	256	3
2005	National public	224	5	263	93	255	2
	California	222	20	258	75	258	5
	Florida	221	4	257	95	250	2
	New York	221	3	267	88	257	9
	Texas	216	6	261	93	243	1
2003	National public	—	—	—	—	—	—
	California	—	—	—	—	—	—
	Florida	—	—	—	—	—	—
	New York	—	—	—	—	—	—
	Texas	—	—	—	—	—	—

— Not available.
Rounds to zero.
‡ Reporting standards not met.
NOTE: The NAEP Reading scale ranges from 0 to 500. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, 2007, 2009, 2011 and 2013 Reading Assessments.

Average scale scores and percentages for mathematics, grade 8 by status as English Language Learner, 3 categories [ELL3], year and jurisdiction: 2013, 2011, 2009, 2007, and 2005

		Yes		No		Formerly ELL	
Year	**Jurisdiction**	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage
2013	National	246	5	287	92	272	3
	California	235	12	284	75	268	13
	Florida	243	5	283	89	275	7
	New York	241	6	285	94	‡	#
2011	National	244	5	287	91	272	3
	California	234	17	283	67	270	16
	Florida	246	5	280	88	273	8
	New York	239	6	283	94	‡	#
2009	National	243	5	286	92	270	3
	California	237	19	280	67	270	14
	Florida	241	5	282	92	270	4
	New York	231	5	285	95	‡	#
2007	National	246	6	284	93	273	2
	California	241	21	279	70	274	9
	Florida	243	5	279	93	265	2
	New York	236	4	282	96	‡	#
2005	National	244	5	281	93	276	1
	California	241	20	275	74	278	5
	Florida	243	5	276	93	257	2
	New York	237	4	282	87	278	9

Rounds to zero.
‡ Reporting standards not met.
NOTE: The NAEP Mathematics scale ranges from 0 to 500. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, 2009, 2011 and 2013 Mathematics Assessments.

NOTE TO Reporting and Dissemination Committee on Embargoed Pre-Release Access to NAEP Reports

The guidelines attached were endorsed by the Reporting and Dissemination Committee in August 2011 as administrative procedures for handling news media requests for embargoed access to NAEP reports to help prepare accurate news stories before the time set for an official release. Recently, application of the guidelines by Governing Board staff has been criticized by an online education news site in New York City, which was denied pre-release access to the 2013 NAEP Reading and Mathematics Report Cards. The Committee will discuss the guidelines and their application at this meeting. It may also discuss whether to prepare guidelines for pre-release access by other persons or organizations that wish to comment on NAEP reports or provide summaries or interpretations for their members or the public.

In addition to the guidelines, this tab includes the report on the Committee discussion in August 2011, information on the recent issue involving the Board's embargo policy, and background materials on embargo practices.

The guidelines pertain only to embargoed pre-release access to NAEP materials by news media personnel and provide for equal treatment of all news organizations, regardless of how their news product is disseminated, whether published, broadcast, or posted on the Internet. Recipients must agree not to make any information public until the time set by the Board for public release.

The guidelines do not apply to education constituency groups, such as a teachers union or school board association, advocacy groups with varying views on education issues, or non-profit think tanks that offer commentary and analysis. At present such groups are not given pre-release access to NAEP reports on the grounds that doing so would, in effect, constitute a general public release because, as a government program, the National Assessment should not pick favorites among them.

Pre-release access to NAEP reports has been given on an embargoed basis to public officials, their staffs, and the organizations representing them that are involved in authorizing, funding, or facilitating the National Assessment. These have included members of Congress, governors, the superintendents of states and urban districts participating in NAEP, senior officials of the U.S. Education Department and the White House. Pre-release briefings have been given to three organizations—the Council of Chief State School Officers (CCSSO) and the National Governors' Association (NGA) for state and national NAEP reports and the Council of the Great City Schools for reports on the Trial Urban District Assessment (TUDA). Embargoed reports are also given to members of the Governing Board and occasionally, upon request, to former members. In addition, pre-release data has been given to NAEP advisory committees and to persons directly involved in preparing the assessments, such as subject-matter experts.

National Assessment Governing Board

News Media Embargo Guidelines

INTRODUCTION

Under law, the National Assessment Governing Board has the responsibility to "plan and execute the initial public release of National Assessment of Educational Progress (NAEP) reports." The NAEP authorizing statute continues that NAEP data "shall not be released prior to the release of [such] reports."

As part of pre-release activities, information is provided to the media in order to facilitate news coverage that reaches the general public. The practice for many years has been to grant access to confidential information to media representatives who have signed an embargo agreement, promising not to print or broadcast news of a report before the scheduled time of release. With the rapid evolution of the media industry bringing new and influential voices through the Internet, more requests for embargoed access are being received from those outside traditional print and broadcast news organizations.

In order for staff to make fair decisions about who should receive embargoed access, objective guidelines are needed. These guidelines establish the criteria and procedures to be used.

FUNCTION AND BENEFIT OF NEWS MEDIA EMBARGOES

Under a longstanding tradition, organizations that release news and research findings to the public have used embargoes as a way to give reporters advance access to the information while retaining control of the timing and nature of their releases. Government officials and agencies, scientific and medical journals, corporate and consumer businesses, and financial institutions often use embargoes, particularly for lengthy or complex information that requires time for thorough review and analysis before news stories are completed.

Embargo agreements can be beneficial to the releasing organization, journalists, and the public that reads the news and can lead to broad-based dissemination and fuller coverage. Embargoed access may achieve the following:

> Give reporters the time to read and analyze reports, to do further research on complex information, to conduct interviews, and to write more complete, nuanced stories before the time set for release. This reduces the chances that a reporter will "dash off" a story quickly and as a result make errors in interpreting data.

Permit news organizations to print or broadcast a story or place it on the Internet as soon as an embargo is lifted, promptly spreading news of the report or research findings to their audiences.

Create interest and buy-in among journalists who are granted access, which may increase coverage. The additional time provided before stories must be written may help journalists appreciate the significance of the information and how newsworthy it is.

RISKS OF EMBARGOES

Embargo breaks may be committed by a news organization or individual seeking to scoop the competition, or they may happen through accident or carelessness.

For most media outlets and individual reporters, the risks of damaging a relationship with a source or attracting negative attention heavily outweigh the possible benefits of violating an embargo agreement. Such cases do happen, but they are rare.

While journalists do not take a formal oath, and need no license, journalistic ethics demand that embargoes—once agreed to—be respected. If a journalist working outside of the traditional media practices ethical journalism, he or she will not knowingly break an embargo.

CRITERIA FOR ACCESS

A requestor must meet one of the criteria below in order to receive embargoed access to NAEP reports:

1) The requestor is an editor, reporter, columnist, or blogger affiliated with a print, broadcast, or online news organization.

Print and broadcast news organizations for which qualifying employees may receive access would include newspapers, magazines, news services, and radio and television news outlets. Some examples: Associated Press, the Bozeman Daily Chronicle, the New York Times, *MSNBC, Fox 5 NY*, the New Yorker, National Review, the Nation, WTOP, Education Week.

Examples of online general-interest news organizations that would receive access: Huffington Post, Daily Kos, the Texas Tribune, the Daily Caller.

Examples of print and online education trade publications and news providers that would receive access: Education Daily, Hechinger Report of Columbia University's Hechinger Institute for Education Journalism, Alexander Russo's This Week in Education, Inking and Thinking on Education by Joanne Jacobs.

2) The requestor is a freelance reporter working on a story for a news organization in one of the categories above.

Requestors may be asked to provide documentation of their employment or freelance assignment**.**

PROCEDURE FOR REQUESTS

Information about the requirements for embargoed access to NAEP reports and embargo agreement forms shall be made available to news media prior to NAEP releases.

A separate agreement form must be signed by each person receiving embargoed information before each release.

DENIAL OF ACCESS

Reporters shall be denied embargoed access to NAEP information if they are not in one of the categories above or refuse to sign the embargo agreement. Those who knowingly break the embargo shall not be granted embargoed access to subsequent NAEP reports for up to two years.

Appeals regarding denial of access shall be determined by the Commissioner of Education Statistics in consultation with the Executive Director of the Governing Board.

National Assessment Governing Board

Reporting and Dissemination Committee

Report for August 5, 2011
EXCERPT

4. Embargo Guidelines for NAEP Releases

The Committee reviewed a set of guidelines, prepared by Governing Board staff, on granting embargoed access to NAEP reports to members of the news media prior to the public release. The guidelines were developed in consultation with NCES, and incorporate Committee views on a previous draft that was discussed at the meeting in May 2011.

Mr. Harris said the guidelines were prompted by major changes in the news business during the past few years with the decline of newspapers, magazines, and television and radio news and the rise of news websites and bloggers on the Internet. The key principle behind the guidelines is that it is in NAEP's interest to grant advance access to its reports to give journalists the time to write better, fuller, more accurate stories. At some point the Board may want to grant advance access to stakeholder organizations, such as teacher unions, or advocacy groups and think-tanks. But it is very difficult to do that without being accused of bias or, on the other hand, giving reports to everyone that requests them, and not having a scheduled release at all.

The proposed embargo guidelines for news media make no distinction among the vehicles used for transmitting the news. They apply equally to print, broadcast, and online media. But they provide that embargoed pre-release access will only be granted to an editor, reporter, columnist, or blogger affiliated with a news organization. The second category that would receive access is freelance reporters working for a news organization, who may be asked to provide documentation of their assignment.

Several Committee members asked how staff would define a news organization as opposed to an advocacy group. Larry Feinberg, of the Board staff, acknowledged that the lines were not always clear but said that a publication or blog that is an offshoot of a policy group or constituency organization, such the professional association of mathematics teachers, would not be regarded as a news organization.

Gov. Sonny Purdue said staff should be very conservative in defining a news organization to avoid having interest groups shape the initial coverage of NAEP results and the conversation around their release. NCES Commissioner Jack Buckley said he felt it is important to be cautious in order to avoid spreading data widely before an official release.

The Committee endorsed the embargo guidelines for implementation by the Executive Director and Board staff as part of the initial public release of NAEP reports. These will be administrative guidelines, not an official policy, and do not require action by the full Governing Board.

CONFIDENTIALITY AGREEMENT:
The Nation's Report Card:
2013 Mathematics and Reading

Under this agreement, you will have access to secure National Center for Education Statistics data that you agree to keep confidential as outlined below.

The data from the report ***2013 Mathematics and Reading*** and statements, commentary, and other materials on the data may not in any way be made public—including print or Internet publication, wire, or broadcast—prior to **Thursday, November 7, at 10 a.m. EST.**

Neither the data nor the access information provided enabling you to view embargoed data and related materials online is to be shared with other individuals or organizations, including on a "hold for release" basis. All conversations about the embargoed data and related materials will be limited to those within your organization who need to be informed for essential work purposes only.

By signing this agreement, you are agreeing that you and your organization will abide by the terms above. Please sign below and send to Shannon Tucker via fax at (703) 299-2424 or email at stucker@reingold.com.

Accepted by:

Signature: _____

Print Name: _____

Email: _____

Organization: _____

Date: _____

The following were denied embargoed pre-release access to materials on the 2013 NAEP Reading and Mathematics Report Cards that were released on November 7, 2013:

1. Mary Tillotson, watchdog.org
2. Kim Greene, Scholastic Instructor and Administrator Magazines
3. Joy Pullman, School Reform News
4. Matt Freidman, Scholastic Math and Science Magazines
5. Tara Welty, Scholastic Instructor Magazines
6. Allison Aubuchon, Foundation for Excellence in Education
7. Carrie Marovich, The Cabinet Report
8. Matt Korobkin, Rodel Foundation
9. Ashley Inman, Education Next
10. Dolly Sullivan, Educate Maine
11. Philissa Cramer, Gotham Schools
12. Kate Schimel, Ed News Colorado

Agency Defends Exclusion of Online News Sites From Early Access to Test Data

blogs.edweek.org /edweek/education_and_the_media/2013/11

By Mark Walsh on November 7, 2013 11:25 PM

|

The federal agency that administers the National Assessment of Educational Progress is standing by its decision to exclude a group of online news organizations from early, embargoed access to test results this week.

The National Assessment Governing Board decided to bar Chalkbeat, Gotham Schools, and Ed News Colorado from the early access provided other news organizations to data and a background briefing on 4th and 8th grade mathematics and reading results.

"Right now reporters being briefed on new #NAEP scores. We're not b/c @GovBoard has inexplicably barred @gothamschools & @Chalkbeat," said a Twitter posting, or tweet, from GothamSchools on Wednesday.

The three sites (GothamSchools, Chalkbeat, and Ed News Colorado) are part of the same organization and are being re-branded under the Chalkbeat banner. They actually learned Monday in an email from a NAGB contractor that they would not be given the embargoed access.

"While the Board understands that these groups may have received access in the past, their current relationship with the Colorado Nonprofit Development Center is considered to be an affiliation with an outside organization and therefore outside of the embargo access policy," the contractor, Reingold Inc., said in the email. "This determination is based the [sic] Governing Board's understanding of the current relationship with the Colorado Nonprofit Development Center with the understanding that Chalkbeat may eventually be an independently funded news outlet."

Elizabeth Green, the editor of the Chalkbeat sites, said in an interview that the Colorado nonprofit center acts solely as the "fiscal sponsor" of the Chalkbeat sites because they have yet to receive their own nonprofit status from the Internal Revenue Service. The center provides back-office services such as human resources, she said.

"Its sole purpose is to incubate non-profit groups," Green said. "It's really not dissimilar from contracting with an outside HR or back-office services firm."

The Chalkbeat sites applied for the same embargoed access given many other print, broadcast, and online news outlets. That includes up to 48 hours advance access to the NAEP results, and a background briefing with officials the day before the expiration of the embargo allows news outlets to publish their stories. For this week's NAEP results, the embargo expired at 10 a.m. Eastern time on Thursday.

"We protested the decision, and I still don't understand their rationale," Green said. "I think it is important to make a distinction between who is an independent news organization and who isn't. But we are. We follow professional standards just like any other news organization."

Cornelia Orr, the executive director of the National Assessment Governing Board, said in an interview that the agency adopted an informal policy about three years ago regarding access to its embargoed data. The policy requires that news organizations be independent, she said.

"We developed a policy that seems to have come right up against Chalkbeat," Orr said. "I'm sympathetic to their being caught here and not having independent status, but that's what it was."

Orr said NAGB has received requests from advocacy groups seeking the same embargoed access that independent news organizations receive. She cited the Education Trust as one example. Such groups are excluded under the

policy. (After some reflection, Orr declined to provide a copy of the informal policy, saying it was an internal document.)

Many federal agencies in Washington have variations of early access to embargoed materials, such as unemployment reports in the U.S. Department of Labor or crop forecasts in the U.S. Department of Agriculture. The pre-publication access is meant to help reporters digest complex data.

"The governing board's first and primary responsibility is to protect the NAEP data and to get the cleanest story about the NAEP data out there," Orr said.

Meanwhile, Orr noted, there have been fast-paced changes in the media in recent years, including many new online news outlets. Among the Web news organizations that did get the embargoed access this week were the Huffington Post and Stateline.org, she said.

Orr also said that in Chalkbeat's case for this NAEP release, there was a fair amount of "11th hour" back and forth before NAGB decided to exclude its sites.

"We didn't want to feel pressured at the end and let Chalkbeat in and not others who had applied earlier," she said. [At least one other online news site, EdSource Today, which focuses on California schools, said late Thursday on a listserv for the Education Writers Association that it had sought access to the embargoed NAEP materials and was turned down. I didn't have the chance to ask Orr about the particulars of that.]

Chalkbeat's Green said her organization got hold of the NAEP data through other sources on Wednesday and prepared stories that were ready to publish as soon as the embargo ended on Thursday morning. (She didn't identify the sources, but said Chalkbeat agreed to observe the embargo time.)

Orr noticed that the Chalkbeat stories had benefited from some alternative early access.

"An embargo is an embargo, and I appreciate that they abided by it," Orr said. "I don't want to come off as unsympathetic. I appreciate their frustration. We will continue to review our policy."

‘Non-traditional’ journalists barred from viewing tax-funded test results early

Posted By Mary C. Tillotson On October 29, 2013 @ 4:00 am

By Mary C. Tillotson | Watchdog.org

A taxpayer-funded research board has refused to share a national study of **American** education with **Watchdog.org** and other “non-traditional” news outlets.

Instead, the **National Assessment Governing Board** is offering early access to the report Tuesday and invitations to discuss it with the authors during a teleconference Wednesday to select media outlets only, including the **Huffington Post**.

The **Nation’s Report Card: 2013 Mathematics and Reading, Grades 4 and 8**, will be available to the public (and those “non-traditional” news outlets) Thursday.

“In a world where we are all bombarded with news 24/7, to discriminate between one form [of journalism] goes against our basic Constitution,” said **Ginger Stanley**, executive director of the **Virginia Press Association**.

Sharon Tucker, communications associate for **Reingold, Inc.**, a private corporation apparently hired to oversee media relations, explained the slight in an email to Watchdog.


[2]

BARRED: Only journalists who qualify under a tax-funded board’s definition can receive early access to this year’s National Assessment of Educational Progress.

“Governing Board policy only permits embargo access for reporters affiliated with traditional news media outlets (e.g. **The Washington Post**, **Chicago Tribune**, **National Geographic**, Huffington Post) that are unassociated with outside organizations (i.e. nonprofit associations, government agencies, academic institutions, for-profit businesses),” the email said.

Tucker’s note offered no other explanation, and she didn’t reply to an email or return repeated phone calls, though a receptionist said Tucker was in the office.

“Some of the most important journalism in the last decade has been done by nonprofit journalists, led by Watchdog.org,” said Watchdog’s senior content editor, **Mark Lisheron**. “**Pro Publica** and **Inside Climate News**, both non-traditional by your definition, have won very traditional **Pulitzer Prizes**.”

Watchdog.org reporters carry state-approved media credentials in more than half of the country's state capitols.

"Watchdog.org vigorously objects to being denied a preview of a taxpayer-funded national study by an agency clinging to a barrier in the world of journalism as useless as the **Berlin Wall**," Lisheron said. "We've earned our place at the press table for our public service."

The **National Assessment Governing Board** doesn't have the resources to handle media requests from every outlet and had to draw the line somewhere, said **Stephaan Harris**, public affairs specialist for the board.

"That's the board's policy," Harris said. "In part, it was a feeling that if you allowed, say, any kind of group or association no matter how small that had any kind of online arm, you'd be opening the floodgates for people claiming themselves as journalists. It's a way to minimize this and manage our embargo process."

The board considered using the number of website hits to determine whether a news agency was qualified to receive early access, but thought it wouldn't be fair to those agencies just below that threshold, Harris said.

"There's really no fair way," he said. "If we say, 'Your website has a minimum of 100,000 original visits, then we can consider you a news organization,' then people who get just under that would cry foul."

Discriminating against certain news agencies violates the **First Amendment**, Stanley said, as does allowing embargoed access at all.

"There should be no distinction between a journalist and a citizen," she said. "All of our laws are citizen laws. Freedom of the press and the First Amendment — the laws of our country, whether they're state laws or federal laws, are citizen laws, and should create an open government for all."

As for scrapping the embargoed early-access policy entirely, Harris said he could bring the idea to the board meeting in December.

The **National Assessment of Educational Progress** began in 1969 and is under **No Child Left Behind.** States only qualify for federal funding if they administer the NAEP math and reading tests every two years to their fourth and eighth graders.

The governing board is selected by the **U.S. Department of Education**, but is officially independent of the department.

Contact Mary C. Tillotson at mtillotson@watchdog.org.

Chalkbeat covers education as a non-profit news site

Greg Toppo, USA TODAY *12:47 p.m. EDT October 22, 2013*

A handful of refugees from print journalism are providing an alternative model for covering state and local schools.

Here's a phrase you don't hear much: newsroom expansion.

That's the goal of a new non-profit news outlet debuting Monday that is gearing up to cover education in-depth in four states, in the process providing an alternative model for local journalism about schools, education policy and education politics.

Created by a handful of refugees from beleaguered — and in a few cases shuttered — print newspapers, the online-only Chalkbeat springs from the unlikely partnership created last January when the New York-based non-profit news site GothamSchools merged with Denver-based EdNews Colorado. Mostly foundation-funded, it gets about one-fifth of its revenue from local sponsorships and job ads for teachers and administrators.

On Monday it's expanding to two more cities with fraught school politics: Indianapolis and Memphis. The network plans to add others as funding from local philanthropists comes calling; it already plans to hire a reporter to cover Nashville schools.

In each bureau, Chalkbeat plans to cover the state legislature and state board of education, as well as the day-to-day developments of schools and districts. They're also demanding that local philanthropy help cover costs as a "public good," much as it would support an art museum or symphony. The network also wants to scale back the role of philanthropy, making each bureau more self-sustaining as it grows.

The outlet's expansion is encouraging news, said Mark Jurkowitz, associate director of the Pew Research Center's Journalism Project. "I think this is further evidence of the growth potential of the non-profit news sector," he said.

A Pew survey last June uncovered 172 non-profit news outlets, most of them tiny startups. What Chalkbeat could represent, Jurkowitz said, is the next step in their evolution as a non-profit essentially franchises its news-gathering model in different cities. "Clearly here is a place where there is a perceived need for coverage of local schools and local school systems that may not be covered as well in the legacy publications," he said.

Media critic Jeff Jarvis, director of the City University of New York's Tow-Knight Center for Entrepreneurial Journalism, sees the effort as a way of "getting back some of the reporting beats that we have lost" as traditional newsrooms shrink. "We know those beats can be businesses now."

516
CONNECT

(

With projected revenues this year of about $2.3 million, Chalkbeat is already bigger than most non-profit news outlets, though a few such as ProPublica and *The Texas Tribune* are quite a bit larger. The Pew survey found that of 77 digital non-profit news outlets willing to disclose revenue, only 14 reported incomes of more than $1 million. Of the 93 willing to reveal staffing levels, most said they had no more than five paid full-time staffers.

When its Indiana and Tennessee bureaus are fully staffed early next year, Chalkbeat will have 22 full-time employees.

That stands in stark contrast to recent trends in newspaper hiring. Though no firm figures exist on the drop in education reporting positions, the American Society of News Editors' most recent annual "newsroom census" found that for the first time since 1978, the overall number of full-time editorial jobs dropped below 40,000. In 2012, newspapers employed about 38,000 reporters, editors and other journalists, nearly one-third fewer than 2000. Just last year, they cut an estimated 2,600 editorial jobs.

Chalkbeat's expansion represents a quiet triumph for its founders, among them the editor and publisher, respectively: Gotham's Elizabeth Green and EdNews' Alan Gottlieb, two journalists bent on social justice and fascinated by education's role in making cities work. Both covered education at big-city newspapers and both have seen their beats slashed by downsizing.

Editor Elizabeth Green, 29, center, talks with Emma Sokoloff-Rubin, 24, from Manhattan, left, and Sarah Darville, 22, from Brooklyn, at Chalkbeat.*(Photo: Jennifer S. Altman for USA TODAY)*

Green, 29, began her journalism career as a student at Montgomery Blair High School, a top-flight school in Silver Spring, Md., where one day in 2000, she recalled, the principal announced over the loudspeaker, "You black and Hispanic students need to get your test scores up!" It was the first time she realized that her school had an achievement gap.

Green began spending her lunch hour interviewing classmates "on the other end of the cafeteria" and wrote up her findings in the school newspaper. Three years later, studying at Harvard, she wrote a nearly 5,000-word expose in *Fifteen Minutes*, the weekend magazine of *The Harvard Crimson*, that took aim at the hollowness of the university's "oft-touted commitment to diversity."

That piece, plus her unabashed wonkishness — Green wrote her senior thesis on Alabama Gov. Bob Riley's failed campaign to make the state's tax code more progressive — landed her a job at *U.S. News and World Report*. In 2007, she began

covering city schools for *The New York Sun,* but the paper imploded in 2008. She and partner Philissa Cramer founded GothamSchools later that year as part of an existing non-profit, OpenPlans.

Gottlieb, 57, came to Denver in 1988 to work at *The Denver Post*, where he covered the city's schools, which at the time were operating under court-ordered busing. Gottlieb became so engrossed in issues of school quality, funding and racial segregation that he finally had to quit the *Post* to write about the issues full time at a local foundation. " He soon moved to EdNews Colorado, founded in January 2008 as a daily blog on education policy in the state legislature.

As with many online news enterprises, Chalkbeat is bristling with veterans: Its Colorado capital editor is Todd Engdahl, a former *Denver Post* city editor who had hired Gottlieb in 1988. Engdahl lost his job during *Post* layoffs in 2007. Chalbeat's Indianapolis bureau chief is Scott Elliott, who's leaving newspapers after 22 years, the last three at the *Indianapolis Star*.

But among the newsroom refugees are a few who may never know what it's like to complain about the dying news business: One of Chalkbeat Colorado's newest hires is Kate Schimel, a 23-year-old reporter who was a one-time intern. "She's never been in a traditional newsroom," Gottlieb said.

News embargo

From Wikipedia, the free encyclopedia

In journalism and public relations, a **news embargo** or **press embargo** is a request by a source that the information or news provided by that source not be published until a certain date or certain conditions have been met. The understanding is that if the embargo is broken by reporting before then, the source will retaliate by restricting access to further information by that journalist or his publication, giving them a long-term disadvantage relative to more cooperative outlets. They are often used by businesses making a product announcement, by medical journals, and by government officials announcing policy initiatives; the media is given advance knowledge of details being held secret so that reports can be prepared to coincide with the announcement date and yet still meet press time. In theory, press embargoes reduce inaccuracy in the reporting of breaking stories by reducing the incentive for journalists to cut corners in hopes of "scooping" the competition.

Embargoes are usually arranged in advance as "gentlemen's agreements." However, sometimes publicists will send embargoed press releases to newsrooms unsolicited in hopes that they will respect the embargo date without having first agreed to do so — the phrase "For Immediate Release" often found at the top of press releases indicates that the information in the release is *not* embargoed.

News organizations sometimes break embargoes and report information before the embargo expires, either accidentally (due to miscommunication in the newsroom) or intentionally (to get the jump on their competitors). Breaking an embargo is typically considered a serious breach of trust and can result in the source barring the offending news outlet from receiving advance information for a long period of time.

News embargoes are one of several ways a source can influence media presentation of the information they provide; others include providing information "on background" or "not for attribution," limiting or providing "access," or even direct government or market intervention against the reporters or media company. (See confidentiality terminology in journalism for a full discussion of these.) The manner in which journalists react to these and other attempts to influence coverage are a matter of journalistic ethics.

Examples of embargoes

- Biweekly press briefings from the International Monetary Fund are typically embargoed until 10:30 a.m. Washington time, 1430 GMT (for synchronised effect on global stock markets).
- Reporters who accompanied U.S. President George W. Bush on a Thanksgiving visit to Iraq in 2003 were embargoed from filing until the President left the country. They were told that, in the interests of security, the trip would be canceled if news broke before its conclusion.[1]
- The Ministry of Defence in the United Kingdom informed a handful of journalism outlets that Prince Harry would be serving in Afghanistan, on condition that the information not be released until the end of his deployment. The information was leaked after about two months, and officials agreed to end the embargo. The prince was immediately removed from the battlefield, reportedly for his safety and that of his fellow soldiers.
- In Canada, Australia and other countries, prior to the release of the budget and other important government announcements, reporters are held in a "lockup" so that they can prepare stories in advance. They are not permitted to file until after the official announcement (for example, after the Minister of Finance rises to deliver the budget speech.) Lockups are particularly aimed at

preventing insider trading on the basis of leaked government announcements.[2][3] A similar lockup is done in the United States when the Federal Reserve Board is preparing to adjust an interest rate.

- *The New York Times* in 2008 prompted suppression of the story of the kidnapping of David Rohde (their reporter) in news outlets and on Wikipedia until his return in 2009. This example, in which the instigator of the embargo is not the source, may be a case of self-censorship instead.

Embargoes on articles in scientific journals

News embargoes are commonly applied on information of health-related news regarding upcoming medical journal articles. All major medical journals, including the *New England Journal of Medicine*, the *Journal of the American Medical Association*, and *The Lancet*, have publication embargoes.

The *JAMA* embargo probably dates back to the editorship of Morris Fishbein, from 1924 to 1949, and holds until 15:00 Central Time on the day before the cover date of the issue. Journalists who agree to not publish (in print, on television, on radio, or via Internet) until that time the information contained in a manuscript to be published by the journal receive advance copies of the journal by mail during the week before publication. For selected articles, press releases and news release videos are also prepared by science writers and released to journalists during that week.[4]

The reasons given for such embargoes are twofold. First, they enable journalists to produce more comprehensive and accurate coverage, as the embargo provides time in which they can research the background to a story and thus publish "backgrounders" along with the story's release. Second, they enable doctors and scientists to receive and to analyze medical studies before the general public does, enabling them to be better informed when called upon to comment or to react by journalists or by patients. However, some object to the medical news embargo system, claiming that it is driven by profit motives on the parts of the medical journals.[4][5]

If you must use embargoes, here's how to do it right

Ivan Oransky[1]

Key words: Embargo; Science journalism; Publishing; Research; Medical news

(1) Founder, Embargo Watch (http://embargowatch.wordpress.com); Adjunct Associate Professor, Arthur L. Carter Journalism Institute, New York University. oransi01@nyu.edu.

DOI: 10.2427/9043

Embargoes - often used by scientific institutions such as medical societies and scientific journals to give access to reporters before material is published - can inspire heated arguments. Some journalists love them, while others say they - along with Ingelfinger Rule, which prohibits pre-publication publicity of results before they appear in a peer-reviewed journal - discourage original reporting [1]. Journals find them helpful in "choreographing" the dance of medical news [2], but some have eschewed them completely [3].

Despite all of this debate, and the fact that embargoes are becoming "less and less practical" [4], in the words of one press officer, they are here to stay, at least for some time.

And as one public relations executive said recently, 'Every manager here has a different set of rules about embargoes' [5].

With that in mind, here are some guidelines for appropriate embargo policies that actually live up to the oft-stated goals of allowing reporters enough time to report stories accurately, while avoiding needless restrictions on the flow of scientific information.

1. **Give a reasonable amount of time.** What's reasonable? That's probably a judgment call, dependent on how complicated the material is, what else is happening in the world, and other factors. Many journals that publish weekly provide embargoed material about five days before publication, which seems like enough time. And I've suggested that 24 hours should be a minimum, even in our millisecond news cycle world. But one thing's for sure: 38 minutes is not long enough [6].
2. **Don't embargo material that's freely available online.** This might appear obvious. But based on the number of journals and scientific conferences that still try to claim that their online accepted papers, corrected proofs [7], and abstracts are embargoed until some time they determine later, it bears repeating. If it's freely available online, it can't be embargoed. End of story.
3. **Give a specific time for your embargoes.** This doesn't come up very often, but saying that a paper is embargoed for a date isn't enough - you have to also say what time, and in what time zone. Otherwise it will lift 26 times [8] for people in 26 different time zones.
4. **Don't ask everyone to agree to an embargo, then let one news outlet go ahead with a story.** I'm as big a fan of original reporting as the next journalist, but I'm not a fan of backroom side deals. If reporters

have agreed to an embargo on your material, don't then give an exclusive to a paper - say, *The New York Times* [9] - while making everyone else wait to publish.

5. **Don't ask for a quid pro quo.** Lots of press officers believe - perhaps with some proof [10] - that embargoes increase the chance something will be covered. But don't make that coverage a condition of your embargo agreement [11]. Reporters may go to your conference for many reasons, including becoming better-informed about a subject, and never write anything that can be pegged to that conference. Quid pro quo is unseemly, not to mention a journalistic no-no.
6. **Be consistent about sanctions and early embargo lifts.** If someone has agreed to your embargo policy, they should get the same sanctions for breaking it as anyone else does, no matter what outlet they work for. Those sanctions should be clearly spelled out in your embargo policy, and you should avoid the temptation to look the other way for repeated "inadvertent" breaks. And don't blame someone for breaking an embargo if he never agreed to embargoes in the first place. Also: Lift the embargo once the material appears online, whether it's an obscure blogger or a major wire service that broke it. If one reason for embargoes is to level the playing field, then keep the playing field level.
7. **Keep the number of cooks in the kitchen to a minimum.** Nowadays, for many journal studies, there are at least two press releases: one from the journal, and one from the researchers' institution. If the research had an industry sponsor, there may be a third. Conferences can get even more complicated, and that's where inadvertent breaks can happen. Do your best to minimize those, and confusion.
8. **If other news is coming out within a day or two of yours, move your embargo so they match.** Let's say you're publishing a study on a particular subject, and your embargo lifts on Thursday at 5 p.m. Eastern. You find out that a competing journal is publishing a study on the same subject at 5 p.m. Eastern on Tuesday. Move yours to Tuesday, and let your press list know. If one reason for embargoes is to allow reporters to write better-informed stories, why insist that they only cover your news if they want to publish at the embargo time [12]? Be flexible. Readers will thank you.
9. **Make sure recipients of your "embargoed" emails have actually agreed to an embargo.** Sending something and marking it "embargoed" doesn't mean it actually is [13]. Just because someone agreed to another institution's embargo policy doesn't mean that she agreed to yours. There's nothing stopping her from writing about the story, and she won't have broken any agreements.
10. **Don't try to restrict with whom reporters can speak.** As a number of embargo policies spell out, part of the reason to give journalists time with material before it's published is so that they can seek outside comment. Requiring that reporters not share the material with anyone before the embargo [14] lifts turns them into stenographers [15]. At the very least, it will make people more cynical about the reasons for embargoes.

References

[1] Kiernan V. Embargoed Science. University of Illinois Press, 2006.
[2] Podolsky SH, Greene JA, Jones DS, "The Evolving Roles of the Medical Journal." N Engl J Med 2012; 366:1457-146. doi: 10.1056/NEJMp1113567
[3] http://www.elifesciences.org/authors-the-media-and-elife/
[4] http://embargowatch.wordpress.com/2013/08/19/astronomy-press-officer-embargoes-are-becoming-less-and-less-practical/
[5] http://www.dontgetcaught.biz/2012/10/embargoes-or-

anything-goes-10-big-myths.html
[6] http://embargowatch.wordpress.com/2013/04/11/new-embargo-record-from-nejm-38-minutes/
[7] http://embargowatch.wordpress.com/2012/10/03/finally-some-honesty-journal-rationale-trumps-logic-for-our-freely-available-but-embargoed-policy/
[8] http://embargowatch.wordpress.com/2010/04/06/an-embargo-that-lifts-every-hour-for-a-day/
[9] http://embargowatch.wordpress.com/2012/02/03/embargoes-and-exclusives-dont-mix-new-york-times-story-about-neuron-reveals-a-problem/
[10] Kiernan V. "Changing Embargoes and the New York Times' Coverage of the Journal of the American Medical Association," Science Communication, March 1998, 19(3):212-221. doi: 10.1177/1075547098019003003
[11] http://embargowatch.wordpress.com/2010/11/10/quid-pro-quo-american-cetacean-society-tells-freelancer-he-can-only-have-a-press-pass-if-its-a-mutually-beneficient-relationship/
[12] http://embargowatch.wordpress.com/2012/11/05/a-shot-in-the-embargo-in-which-i-politely-call-bullshit-on-the-journal-pediatrics/
[13] http://embargowatch.wordpress.com/2010/10/21/youre-doing-it-wrong-sending-material-and-calling-it-embargoed-before-an-agreement-doesnt-make-it-so/
[14] http://healthjournalism.org/blog/2011/06/fda-creates-embargo-policy-in-response-to-reporter-concerns/
[15] http://embargowatch.wordpress.com/2012/09/21/stenographers-anyone-gmo-rat-study-co-sponsor-engineered-embargo-to-prevent-scrutiny/

*

Governing Board and Committee Input in NAEP Results

The Reporting and Dissemination Committee is continuing an ongoing discussion of its role in the reporting, release, and dissemination of National Assessment of Educational Progress (NAEP) results. The Committee desires more input at the beginning, or conception, phase of report development, rather than solely providing feedback on a late-stage draft report or web site. The goal is to have input at a "big picture" level rather than provide detailed edits. Also, being mindful of the changing media landscape and the need to make NAEP relevant to diverse audiences, the Committee is exploring the development of additional focused reports on important aspects of NAEP data that can extend the impact of NAEP results beyond their initial release.

The Committee has expressed interest in ensuring that NAGB can impact NAEP reporting while preserving the legal responsibilities of the Governing Board, which sets policy for NAEP, and the National Center for Education Statistics (NCES), which assesses the students, analyzes the data, and uses the findings to prepare NAEP reports. The Governing Board's NAEP reporting, release, and dissemination policy (in full below), adopted in 2006, was used as a starting point for this discussion.

After the Board meeting in August 2013, Committee Chair Andrés Alonso requested Board and NCES staff to begin collaboration on possible ideas to achieve the Committee's goals for discussion at the December Board meeting. Important context for this discussion includes the fact that data from NAEP Report Cards are transitioning to being released mostly online through an interactive site as opposed to printed reports. In advance of report public releases, members and staff would see preliminary results through a preview of the interactive site and a brief printed summary. Additionally, the process for allowing earlier and higher-level input is complicated by the six-month reporting window for Report Cards in subjects like mathematics and reading. The ability to see data in a consumable form before a site preview would be a challenge. Preliminary discussion ideas and suggestions are listed below.

- **Pre-Data Discussions:** At the Board's March meeting, Committee members can start discussion on assessments being undertaken for that calendar year. So for 2014, this would include NAEP Civics, U.S. History, Geography, and Technology and Engineering Literacy. Though testing is yet to be completed and there will be no data at that point, the Committee can discuss what types of data, trends, comparisons, etc., should be included and highlighted on the NAEP report site. These views can inform visioning meetings conducted by NCES and its NAEP contractors where data will be discussed and report structure determined.

- **Singling Out Topics:** Committee members can suggest topics within a subject they think the public might be especially interested in and then the website can highlight that in some way. In U.S. history, for example, if Committee members believe that topics such as the Civil War or the Civil Rights Movement would have a wide appeal, the Nation's Report Card website can give prominence to the test items, response rates, subgroup performance, etc., for that topic.

- **Guiding Questions:** Committee members can suggest ideas for the main questions on the interactive NAEP website through which performance summaries and charts and tables are structured. The site for NAEP 2013 Reading and Mathematics Report Card, for example, has three big-picture questions such, as "Are Students Making Progress?," designed to tell a story through the findings.

- **Ideas for NAEP Website Graphics:** Committee members can suggest general ideas for potential trends, comparisons, etc., that would be make for a good chart or table. The purpose would be to highlight "hidden gems", trends or patterns that normally are not covered in the media but to which NAEP should bring more attention.

National Assessment Governing Board
Reporting, Release, and Dissemination of NAEP Results
Policy Statement

Adopted: August 4, 2006

The Nation's Report Card ***informs the public about the academic achievement of elementary and secondary students in the United States. Report cards communicate the findings of the National Assessment of Educational Progress (NAEP), the only continuing and nationally representative measure of achievement in various subjects over time.*** **The Nation's Report Card** ***compares performance among states, urban districts, public and private schools, and student demographic groups.***

Introduction

NAEP collects data through representative-sample surveys and reports fair and accurate information on academic achievement to the American public. By law (P.L. 107-110, as amended by P.L. 107-279), NAEP is administered by the Commissioner of the National Center for Education Statistics (NCES) under policy set by the National Assessment Governing Board ("the Governing Board"), a bipartisan, independent policymaking body.

According to the statute, the Governing Board shall exercise "independent judgment, free from inappropriate influences and special interests" and in the exercise of its responsibilities, "shall be independent of the Secretary and the other offices and officers of the Department [of Education]." Among the responsibilities specifically delegated to the Governing Board are: (1) "develop guidelines for reporting and disseminating [NAEP] results"; (2) "take appropriate actions needed to improve the form, content, use, and reporting of [NAEP] results"; and (3) "plan and execute the initial public release of [NAEP] reports."

To carry out these responsibilities, the Governing Board hereby adopts policy principles and guidelines for the reporting, release, and dissemination of The Nation's Report Card.

As outlined in the appendix, this policy defines The Nation's Report Card as, and applies to, the initial reporting of NAEP results from national, state, and trial urban district assessments (TUDA), and to other special reports or studies authorized by the National Assessment Governing Board, including printed reports and the initial release Web site.

Delineation of NAEP Reporting, Release, and Dissemination Responsibilities

The NCES Commissioner, under Governing Board policy guidance, is responsible for administering the assessment, ensuring the technical soundness and accuracy of all released data, preparing NAEP reports, and presenting NAEP results.

In addition to setting policy, Governing Board is responsible for ensuring policy compliance of Governing Board-authorized NAEP reports, determining their respective dates of release, and planning and executing the initial public release of NAEP results.

Part I: Report Preparation and Content

Policy Principles

1. The primary means for the initial public release of NAEP results shall be a printed summary report, known as The Nation's Report Card, accompanied by a separate, dedicated Web site – http://nationsreportcard.gov.

2. The primary audience for The Nation's Report Card is the American public.

a. All reports shall be written in language appropriate for an audience of the interested general public, the majority of whom are unlikely to have a technical understanding of education statistics or assessment.

3. The Nation's Report Card shall report data objectively, accurately, clearly, and fairly, in accordance with NCES data quality standards. Results shall be insulated from ideological and other special interests.

a. The Nation's Report Card shall include straightforward presentations of data. Reports may suggest correlations, but should not conclude cause-and-effect relationships. Any interpretation of results must be strongly supported by NAEP data.

b. The Nation's Report Card and its Web site may include references and links to the National Assessment Governing Board Web site, NCES Web site, and the NAEP Validity Studies Panel. Non-NAEP materials and links to non-NAEP resources shall not be included in initial release documents, with the exception of relevant federal and state government information, such as NCES surveys and other district, state, national, or international testing programs.

c. To improve public understanding of results, The Nation's Report Card should contain information about Governing Board-approved NAEP contextual variables and subject-specific background information—as outlined in the Background Information Framework for the National Assessment of Educational Progress (adopted by the National Assessment Governing Board, 8/1/03)—when available and reliable. Reports may also contain other contextual information from trustworthy sources outside of the NAEP program, such as expenditures per pupil, student/teacher ratios, and student enrollment.

4. In accordance with the law, The Nation's Report Card shall include results for the nation; states and school districts, when collected in conjunction with specific NAEP programs,; and school types, disaggregated by subgroup whenever reliable. Subgroup results shall be prominently positioned to facilitate public review but shall not be used to adjust findings.

a. Disaggregated subgroup data should be accompanied by information about demographic changes in the student population assessed.

b. Results for states and school districts may be presented in alphabetical or rank order, accompanied by appropriate language to make the public aware of any data comparison limitations.

c. Data shall be publicly released on inclusion and accommodation rates for all NAEP samples, including national, state, district, and school type. Results for students with disabilities and English language learners shall be presented separately.

5. The Nation's Report Card shall report results by Governing Board-adopted achievement levels, average scale scores, and percentile distributions. Trend information shall be an important part of reports unless comparable and reliable data are not available.

a. Reports shall contain clear explanations of achievement levels, including item maps and sample test questions and answers to illustrate what students in each grade assessed should know and be able to do at each achievement level.

6. All NAEP data determined by the NCES Commissioner to be valid and reliable shall be made available on the World Wide Web at the time of initial public release, except for data from limited special purpose samples and pilot studies. A separate, dedicated Web site aimed at a broad public audience – http://nationsreportcard.gov – shall be utilized for initial public releases.

a. All released NAEP data shall be subject to NCES quality control procedures to ensure accuracy and completeness.

b. At least one block of released NAEP questions shall be posted on the World Wide Web for each subject and grade for which results have been collected.

c. Concise information on test content, methodology, performance standards, and scoring shall be included in all NAEP reports. More extensive material on these topics should be readily accessible on the World Wide Web.

7. Results of special studies authorized by the Governing Board will be reported after careful review of information quality and statistical validity. These shall be treated as initial public releases of The Nation's Report Card, and shall be subject to NCES quality control procedures and Governing Board policies.

8. The Governing Board shall adopt general guidelines to inform the development of The Nation's Report Card and its Web site, and may set additional specifications for particular reports.

9. The Governing Board shall review the format and content of initial releases, including Web pages, to ensure compliance with Governing Board policy.

a. The Nation's Report Card shall contain a description of the policymaking roles and responsibilities of the National Assessment Governing Board, including a list of current Governing Board members, their affiliations, and regional locations.

Part II: Public Release of NAEP Results

Policy Principles

1. Release activities shall be planned and executed by the National Assessment Governing Board. The Governing Board shall determine the release date, time, embargo policies, and manner of release for The Nation's Report Card, as covered by this policy.

a. After the Governing Board has approved the final draft of The Nation's Report Card, including the pages that will be made available through the initial release Web site, the Chairman of the Reporting and Dissemination Committee, on behalf of the Governing Board, shall determine the date of the initial public release, in consultation with the Chairman and Executive Director of the National Assessment Governing Board and the NCES Commissioner.

b. The initial release shall be completed within 30 days of approval of the final draft of The Nation's Report Card. In setting that release date, attention will be paid to balancing the priorities of an expeditious release with provision for adequate planning time, given the scheduling circumstances of the various parties involved.

c. Prior to the initial public release, NAEP results may be provided on an embargoed basis to federal, state, and TUDA-district officials and members of the press.

2. The Governing Board shall be responsible for organizing and conducting the release event and related activities.

a. A release plan shall be adopted by the Governing Board for each report. Elements of the plan may include issuance of a press release, a press conference and/or Web-based announcement, distribution of summary findings and graphics, time period for the initial public release phase of http://nationsreportcard.gov, and other related activities.

b. The official press release announcing NAEP results shall be issued by the Governing Board. Accompanying statements from the Governing Board's Executive Director or Governing Board members may also be issued.

c. At the press conference or other event for release of NAEP results, the NCES Commissioner or his/her designee shall present major data findings, accompanied by a written statement. The National Assessment Governing Board shall select members to provide individual commentary on the meaning of results. In addition, the Governing Board may invite other officials or experts to comment on the significance of the results in accordance with the approved release plan.

d. At press conferences, questions from the audience shall be limited to accredited members of the media. At other public release events, the Governing Board shall determine who may attend and ask questions or comment.

3. The Nation's Report Card shall seek to encourage wide public attention to NAEP results and clear understanding of their meaning and significance.

a. Video materials may be prepared to accompany the release. These shall be clearly identified as having been provided by the Governing Board or NCES of the U.S. Department of Education. The video materials may only contain sound bites, background

footage, and other information for journalists to develop their own stories.

4. Release procedures shall underscore the credibility of The Nation's Report Card and encourage the participation of schools, school districts, and states in NAEP.

a. NAEP data in statements distributed at The Nation's Report Card initial public release events shall be checked for accuracy by NCES.

5. The Nation's Report Card releases shall be clearly separated from any ideological or other special interests.

a. Activities related to the initial public release of The Nation's Report Card shall not be used to disseminate any materials unrelated to NAEP. No materials of any kind may be distributed at an initial release event without the prior approval of the Governing Board.

6. The National Assessment Governing Board will cooperate with the NCES Commissioner in the release of technical reports, working papers, and secondary analyses not covered by the policy.

7. The Governing Board will develop a reporting schedule each year for upcoming NAEP assessments based on data review and report production plans that are provided and updated by NCES.

Part III: Dissemination and Outreach

Policy Principles

1. Information from The Nation's Report Card shall be disseminated through the media, the World Wide Web, and special publications and materials. Efforts shall be made to develop widespread public awareness of NAEP data and their meaning and of the value of The Nation's Report Card to the nation and participating jurisdictions.

a. NAEP results shall be available in both printed and electronic form, including on The Nation's Report Card Web site, at the scheduled time of release and in the permanent record.

b. To build public awareness of The Nation's Report Card, the home page of the initial release Web site shall remain on-line and include links to previous releases. This homepage shall link to respective pages found on the NAEP Web site.

2. To build understanding of The Nation's Report Card and the data it reports, other information about NAEP may be disseminated at the time of the initial release and on a continuing basis.

a. Informational materials accompanying results shall explain the mission and value of The Nation's Report Card in clear and compelling terms.

3. The Nation's Report Card and supplementary NAEP materials shall be made available through a wide network of education, business, labor, civic, and other interested groups and to policy makers and practitioners at all levels of education and government.

a. The Nation's Report Card shall be distributed promptly to governors and chief state school officers, as well as to superintendents of TUDA districts. The reports shall be posted on the World Wide Web immediately at the time of initial release, with printed copies available to the public upon request.

b. Notification of upcoming releases shall be widely disseminated. Schools and school districts participating in NAEP samples shall be provided with information on how to access reports electronically and obtain printed copies upon release.

c. NCES and Governing Board staff shall encourage national and state organizations that are interested in education to disseminate NAEP results to their members.

d. The NCES Commissioner and staff, Governing Board members and staff, and NAEP State Coordinators are encouraged to increase awareness and understanding of NAEP among the public, educators, and government officials. They are encouraged to speak about the NAEP program to a variety of audiences; at meetings and conferences of national, state, and local organizations; on radio and television; and to writers for magazines and newspapers and other members of the media.

e. Talking points on key data findings shall be developed for each release and distributed to Governing Board members.

4. A variety of materials shall be developed, appropriate to various audiences, to carry out NAEP dissemination. Key audiences for these materials shall include the interested general public, policymakers, teachers, administrators, and parents.

5. Detailed data on cognitive results, Governing Board-approved contextual variables, and subject-specific background information (as outlined in Part I, Policy Principle 3, Item C) shall be made readily available through the World Wide Web to all those wishing to analyze NAEP findings, subject to privacy restrictions. Additional restricted data shall be available for scholarly research, subject to NCES licensing procedures.

a. The limitations on interpretations, conclusions, and recommendations in official NAEP reports (as outlined in Part I, Policy Principle 3) shall apply fully to any materials disseminated as part of the NAEP program by NCES and the Governing Board.

b. Researchers receiving secondary analysis grants from NCES may analyze data and provide commentary. Their reports may be disseminated by NCES if they meet NCES standards.

Appendix: NAEP Initial Release Reporting Covered by this Policy

The Nation's Report Card

The primary means for the initial public release of NAEP results shall be a summary report in each subject, known as The Nation's Report Card™ and intended for the interested general public. The reports shall be made available in both print and electronic (Web-based) form. These reports shall present key findings and composite and disaggregated results. The printed reports shall be relatively brief, and written in a clear, jargon-free style with charts, tables, and graphics that are understandable and attractive. Data tables may be included in an appendix, either bound into the report or printed separately. This format shall be used to report key results for the nation and the states and of NAEP Trial Urban District Assessments.

A separate, dedicated Web site for the initial release of NAEP results shall be focused on a broad public audience, including less sophisticated users of the technology. The URL – http://nationsreportcard.gov – should be readily located via Internet search engines. Key NAEP findings will be available, clearly organized and prioritized. World Wide Web pages shall provide key findings, including composite and disaggregated results, as well as access to more extensive data sets.

Individual State and School District Reports

Relatively brief reports of key results shall be prepared for individual states, as well as for TUDA-participating school districts. All reports shall contain composite and disaggregated data, and may include an appendix with data tables.

Special Studies and Reports

Special studies and reports authorized by the National Assessment Governing Board and based on NAEP data collections will focus on specific topics of public interest and educational significance. They are aimed at policymakers and interested members of the public. They may include newly released data as well as data previously released that are analyzed to address issues identified by the Governing Board.

THE NATION'S REPORT CARD

2013 MATHEMATICS AND READING, GRADES 4 AND 8 RELEASE WEBINAR

EVENT DEBRIEF

OVERVIEW

On November 7, 2013, the National Assessment Governing Board coordinated a live webinar to release results of the 2013 National Assessment of Educational Progress (NAEP) *Mathematics and Reading, Grades 4 and 8.* Panelists included:

- **Jack Buckley,** Commissioner, National Center for Education Statistics
- **Honorable Leticia Van de Putte,** Texas State Senator; Member, National Assessment Governing Board
- **William D. Waidelich,** Executive Director, Association for Middle Level Education
- **Cornelia Orr,** Executive Director, National Assessment Governing Board *(moderator)*

The report card results were announced at the webinar, as was the new, interactive Nation's Report Card website. As a demonstration showed, the new online format makes NAEP data more accessible and visually appealing in many ways, for example, by positioning mathematics and reading results side by side for the first time, and featuring videos that guide visitors on how to easily explore and interpret results. As was also explained, the site organizes important findings and trends through big-picture questions to reveal how students are performing by race and ethnicity, gender, income level, and additional contextual factors.

For the first time, the webinar extended the question and answer session, so the event lasted 90 minutes instead of 75 minutes. Thirty-six questions were submitted before the webinar.

WEBINAR ATTENDANCE

This release saw a record number of RSVPs, with 741 registrants.

- In attendance: 433
- Unique organizations represented: 298 (internal staff and contractors excluded)
- State departments of education had a strong turnout; more than one-third of attendees were from public schools or education agencies
- Education organizations attended in high numbers; 1 in 5 attendees were from such groups
- First-time attendees comprised 61 percent

A post-webinar survey drew responses from 118 attendees, in which they were asked questions about their satisfaction with the event and suggestions for future events. The responses were overwhelmingly positive:

- 95 percent found the information from the event "very relevant" or "relevant" to their work
- 83 percent of respondents said they would be willing to help spread the word for future release events
- Webinar event attendees commented favorably on both the webinar and the new online report format. Comments included:
    - "Very helpful - especially walk through how to use website resources." —*staff member from Massachusetts Business Alliance for Education*
    - "Glad to see the Nation's Report Card website updated. Looks like it will be aesthetic and easy to use!" — *staff member from Mississippi State University*
    - "This was a very helpful web event as it walked through the available information and means to access it, as well as offering a variety of perspectives and opportunity for Q&A." — *staff member from North Dakota Department of Public Instruction*

For the first time, the Board is conducting user surveys with media and congressional staff to gather feedback on the usability and accessibility of the new online Nation's Report Card format. Survey results will be presented to committee members at the Board meeting.

TRADITIONAL MEDIA COVERAGE

A full overview of release media coverage will be presented to committee members at the Board meeting. Initial highlights include:

Embargo Access Requests and Media Conference Call

- A record-breaking 164 reporters registered for access to embargoed report card data.
- 55 reporters attended the embargoed media conference call on November 6.
- 42 of the news outlets on the call published articles.
- There was one media embargo break, by La Opinion reporter Maria Pena, who posted about the results to her personal Twitter account. She removed the post when asked to.

U.S. Secretary of Education Arne Duncan hosted a separate national conference call with journalists prior to the release date, which further promoted NAEP data.

Media Coverage

Within 24 hours of the release event, 15 national news outlets published 21 original stories about the math and reading report in print or online.

National Coverage

- Associated Press
- Atlantic
- Bloomberg
- Education Week (two stories, and two blog posts)
- The Educated Reporter (Education Writers Association blog)

- The Huffington Post
- Mother Jones
- The New York Times
- Politico Morning Education (two stories)
- Stateline (Pew Charitable Trusts news service)
- U.S. News & World Report (two stories)
- USA Today
- The Wall Street Journal
- The Washington Post (three stories)

Local Coverage

- 173 local news organizations published original stories about the report in print or online.
- There were 360 local broadcasts about the release on television and radio.

Wire Distribution

In addition to the original news stories developed by the Associated Press, three wire services distributed news of the results:

- AP Top Headline
- Federal News Service
- UWire, a university newswire

The news release was posted by 216 websites within 48 hours of the release event.

Eighteen news outlets used infographics citing NCES data, including the Huffington Post and Washington Post.

SOCIAL MEDIA COVERAGE

This release saw a burst of social media activity about NAEP and the Governing Board larger than any previous release, including:

- Approximately 1,300 conversations about the report during the November 7 webinar.
- More than 4,200 conversations on November 7, totaling more than 10,000,000 potential impressions.

Ninety percent of the online conversation took place on Twitter:

- NAEP was the top trending topic in the DC-area on Twitter during the webinar event
- Governing Board messages were retweeted 25 times on November 7, reaching a potential 333,426 people.
- In total, 18 stakeholders promoted the release event on Twitter
- The most influential tweets (based on reach and engagement) came from the Department of Education, NAEP/NCES, the Governing Board, Education Week, and Democrats for Education Reform.

The following link provides a broad representation of the online conversation leading up to, during, and after the release: http://storify.com/GovBoard/the-nation-s-report-card-2013-mathematics-and-read/.

The Nation's Report Card:
2013 Mathematics and Reading, Grades 4 and 8
24-Hour News Media Coverage
November 15, 2013

Within one day of the release event, news organizations published **195 original stories** about the 2013 Mathematics and Reading report: **23** in national outlets and **172** in local outlets.

The headlines below link directly to the respective articles online. Should an outlet change the article URL and the link become broken, we are happy to provide the full text of the article.

National Outlets

Not good enough: Math, reading scores up slightly
Associated Press—Kimberly Hefling
Published November 7, 2013

American Math and Reading Skills Are Slowly Getting Better
The Atlantic—Julia Ryan
Published November 7, 2013, 10:36 a.m.

U.S. Schoolchildren Show Gains in Nation's Report Card Tests
Bloomberg—Oliver Staley
Published November 7, 2013, 10 a.m.

How States Present--and Spin--NAEP Scores for the Public
Education Week—Andrew Ujifusa
Published November 7, 2013, 3:56 p.m.

U.S. Math, Reading Achievement Edges Up, But Gaps Remain
Education Week—Catherine Gewertz (subscription required; full story can be read on Hechinger Report)
Published November 7, 2013

Arne Duncan 'Encouraged' By NAEP Results
Politics K-12 (Education Week blog)—Alyson Klein
Published November 7, 2013, 10 a.m.

NAEP Score Boosts: Was It the Teaching?
Teacher Beat (Education Week blog)—Stephen Sawchuk
Published November 7, 2013, 4:57 p.m.

The Nation's Report Card: A Slow Climb Up a Steep Hill
The Educated Reporter (Education Writers Association blog)—Emily Richmond
Published November 7, 2013

National Test Scores Show Slight Math, Reading Increases for American Students
The Huffington Post—Joy Resmovits
Published November 7, 2013, 10 a.m; updated 4:40 p.m.

New NAEP Scores Show Continued Improvement in American Schools
Mother Jones—Kevin Drum
Published November 7, 2013, 12:32 p.m.

U.S. Reading and Math Scores Show Slight Gains
The New York Times—Motoko Rich
Published November 7, 2013

It's time to fight - Education world locked in schoolyard brawl - Louisiana voucher program analysis out
Politico Morning Education (10th item)—Caitlin Emma
Published November 8, 2013, 10:01 a.m.

National Report Card Day - Watch for Louisiana voucher data today - Education Department gets an earful - Community colleges get a boost
Politico Morning Education—Libby A. Nelson
Published November 7, 2013, 10:02 a.m.

Reading, Math Scores Inch Up
Stateline.org (Pew Charitable Trusts news service)—Adrienne Lu
Published November 7, 2013

Education Reform May Leave High-Performing Students Behind
U.S. News & World Report—Allie Bidwell
Published November 8, 2013

Racial Achievement Gaps Remain Largely Unchanged, Despite Higher Test Scores
U.S. News & World Report—Allie Bidwell
Published November 7, 2013

Tennessee and D.C. lead education reform: Column
USA Today—Richard Whitmire
Published November 7, 2013, 4:52 p.m.

U.S. Students Make Slight Progress on Test Scores
The Wall Street Journal—Stephanie Banchero
Published November 7, 2013; updated 7:41 p.m.

U.S. students show incremental progress on national test
The Washington Post—Lindsey Layton
Published November 7, 2013

D.C. posts significant gains on national test, outpacing nearly every state
The Washington Post—Emma Brown
Published November 7, 2013

Amid testing gains, D.C. students exhibit achievement gaps
The Washington Post—Emma Brown
Published November 8, 2013

Local Outlets

Race gap expands on some Ohio academic tests; scores become fodder for position statements
Akron Beacon Journal—Doug Livingston
Published November 9, 2013, 12:58 a.m.

Report card: State fails to improve
Albuquerque Journal—Mike Bush
Published November 7, 2013; updated November 8, 2013, 12:05 a.m.

Michigan lagging behind on 'Nation's Report Card' in mathematics, elementary reading
All Michigan/MLive.com—Brian Smith
Published November 7, 2013, 12:48 p.m.; updated 1:07 p.m.

State 8th-graders gain on U.S. reading test
Arkansas Democrat-Gazette—Cynthia Howell
Published November 8, 2013

Arkansas 4th- and 8th-grade students score below national average
Arkansas Democrat-Gazette—Christina Huynh
Published November 7, 2013, 4:02 p.m.

New Jersey students outperform other states in NAEP results
Asbury Park Press—Gina Columbus
Published November 7, 2013

US report: NH students score well in math, reading
Associated Press—Rik Stevens
Published November 7, 2013

Tenn. students lead nation in improvement
Associated Press—Lucas L. Johnson II
Published November 7, 2013

Kansas NAEP scores in math, reading steady in 2013, still top national average
Associated Press—No Author Listed
Published November 7, 2013

NAEP results released; Wyoming beats national avg
Associated Press—No Author Listed
Published November 8, 2013

Wis. black students' reading scores rank low in US
Associated Press—No Author Listed
Published November 7, 2013, 8:36 p.m.

Report Card: Nevada students lag in math, reading
Associated Press—Sandra Chereb
Updated November 7, 2013, 3:46 p.m.

South Dakota Student Math, Reading Scores Steady
Associated Press—No Author Listed
Published November 7, 2013, 5:43 p.m.

Fla. math and reading scores rise slightly
Associated Press—No Author Listed
Published November 7, 2013

Glance: 4th and 8th grade math and reading scores
Associated Press—No Author Listed
Published November 7, 2013, 7:03 a.m.

Report card: Oregon's math, reading scores steady
Associated Press—Gosia Wozniacka
Published November 8, 2013, midnight

Duncan praises Hawaii's math, reading test scores
Associated Press—Jennifer Sinco Kelleher
Updated November 7, 2013, 1:50 p.m.

Report: Maine students score well in math, reading
Associated Press—No Author Listed
Published November 7, 2013, 11:45 a.m.

US report card: Missouri's scores hold steady
Associated Press—No Author Listed
Published November 7, 2013, 11:54 a.m.

Kansas NAEP Scores Flat in 2013
Associated Press—No Author Listed
Published November 7, 2013, 11:47 a.m.

New North Carolina schools report card measures life skills
Associated Press—Emery P. Dalesio
Published November 7, 2013

Dayton, Minn. education commissioner hail progress of black students on national math tests
Associated Press—Patrick Condon
Published November 7, 2013, 2:45 p.m.

Pence: Data shows Indiana students second in gains
Associated Press—No Author Listed
Published November 7, 2013

NJ students score highly on national test
Associated Press—Geoff Mulvhill
Published November 7, 2013

Wash. scores up on national reading, math test
Associated Press—Donna Gordon Blankinship
Published November 7, 2013, 12:36 p.m.

Report: NY's 4th-, 8th-graders scores up slightly
Associated Press—Carolyn Thompson
Published November 7, 2013; updated 3:21 p.m.

Arizona students still lag nation in math, reading
The Republic—Mary Beth Faller
Published November 7, 2013, 11:32 p.m.

New NAEP scores released: Georgia shows progress
The Atlanta Journal-Constitution—Maureen Downey
Published November 7, 2013, 10:14 a.m.

Reading and math scores rise in Georgia
The Atlanta Journal-Constitution—Wayne Washington
Published November 7, 2013, 9:59 a.m.

Texas reading scores lag for all students
Austin American-Statesman—Melissa B. Taboada
Published November 7, 2013, 6:09 p.m.

Report shows Wisconsin black students scoring lowest in nation
The Badger Herald (Madison, Wis.)—Joel Witt
Published November 7, 2013 11:30 a.m.

Maryland students show no significant gains on national tests
The Baltimore Sun—Liz Bowie
Published November 7, 2013, 6:32 p.m.

Mass. students score well on national assessment exam
The Boston Globe—Jasper Craven
Published November 7, 2013

National report: Vermont test scores high, but flat
Brattleboro Reformer—Howard Weiss-Tisman
Published November 8, 2013, 3 a.m.

New York students hover near national average in math, reading
Capital (Albany, N.Y.)—Jessica Bakeman
Published November 7, 2013, 11:26 a.m.

In the News: Modest gains seen in NAEP results
Catalyst Chicago—Cassandra West
Published November 8, 2013

W.Va. students improve slightly in reading, math
The Charleston Gazette—Mackenzie Mays
Published November 7, 2013

New N.C. exams paint bleak picture of skills in state, CMS
The Charlotte Observer—Ann Doss Helms
Published November 8, 2013

US 'report card' for 2013: Student achievement creeps upward
The Christian Science Monitor—Amanda Paulson
Published November 7, 2013

State math, reading scores flat; large percentage below basic achievement
Clarion-Ledger—Sam R. Hall
Published November 7, 2013, 12:32 p.m.

Nation's Report Card: Only a Third of 8th Graders Can Read, Compute At Grade Level
CNSNews.com (Cybercast News Service)—Barbara Hollingsworth
Published November 7, 2013, 1:04 p.m.

'Nation's Report Card' releases state data
Columbia Daily Tribune (Mo.)—Tribune Staff
Published November 7, 2013, 2 p.m.

Ohio students' reading, math scores not improving
The Columbus Dispatch—Catherine Candisky
Published November 8, 2013, 6:42 a.m.

2013 Nation's Report Card: Tennessee shows nation's best education gains
The Commercial Appeal—Richard Locker, Zack McMillin, Jane Roberts, Grant Smith
Published November 7, 2013, 9:37 a.m.; updated 7:24 p.m.

N.H. students outperform national average in reading, math
Concord Monitor—Kathleen Ronayne
Published November 7, 2013; published in print on November 8, 2013

Nation's math and reading scores show progress
Education Matters (Connecticut Post blog)—Linda Conner Lambeck
Published November 7, 2013

Calif. student scores in reading, math climb, but mixed overall
Daily Democrat (Woodland, Calif.)—Barbara Jones
Published November 8, 2013, 12:13 a.m.

State test scores remain flat
Daily Journal (Tupelo, Miss.)—Chris Kieffer
Published November 8, 2013

Texas Hispanic students lag in "Nation's Report Card"
The Dallas Morning News—Terrence Stutz
Published November 8, 2013

D.C.'s Fourth And Eighth-Graders Outpace Nearly Every State in Nation's Report Card
DCist—Matt Cohen
Published November 7, 2013, 12:15 p.m.

Delaware's scores on national student tests show progress
The News Journal—Matthew Albright
Published November 8, 2013

NAEP scores rise for Colorado 4th-graders, 8th graders lose ground
The Denver Post—No Author Listed
Published November 7, 2013, 4:29 p.m.

Utah schools improving in reading, achievement gap, new report shows
Deseret News—Benjamin Wood
Published November 7, 2013, 12:55 p.m.

U.S. students make small gains in reading and mathematics
Deseret News—Celia R. Baker
Published November 7, 2013, 8:35 a.m.; updated 2:11 p.m.

Michigan math exam scores trail U.S. average for 4th-, 8th-graders
The Detroit News—Jennifer Chambers
Published November 7, 2013, 11:56 p.m.

Nation's report card results show some Iowa education gains, needs
The Gazette—Rod Boshart
Published November 7, 2013, 12:30 p.m.

20 Years Later, U.S. Students Making Big Academic Gains
EdMedia Commons (National Education Writers Association)—Mikhail Zinshteyn
Published November 7, 2013, 11 a.m.

Colorado middle schoolers fall short on national report card
EdNews Colorado—Kate Schimel and Sarah Darville
Published November 7, 2013

Rise & Shine: Colorado reassesses school finance after Amendment 66's defeat
EdNews Colorado—Kate Schimel
Published November 7, 2013

Tennessee education gains top the list
Elizabethton Star—Kayla Carter
Published November 8, 2013, 11 a.m.

Tennessee students lead nation in academic growth
WHBQ—Matt Gerien
Published November 7, 2013, 4:19 p.m.; updated November 14, 2013, 5:55 p.m.

A few states, but not N.Y., see big gains on 'nation's report card'
Gotham Schools—Sarah Darville
Published November 7, 2013, 10 a.m.

'Nation's Report Card': Connecticut Students Strong in Reading, Less So In Math
The Hartford Courant—Kathleen Megan
Published November 7, 2013, 11:54 a.m.

Hawaii's math scores add up to first-ever ranking
HawaiiNewsNow (Kaimuki, Oahu)—Jim Mendoza
Published November 7, 2013, 10:49 p.m.; updated November 8, 2013, 12:14 a.m.

Students improve in math, reading
Hawaii Tribune-Herald—Erin Miller
Published November 8, 2013, 12:05 a.m.; updated 12:06 a.m.

Local schools rank 12th in new NC assessment
Hendersonville Lightning—Bill Moss
Published November 7, 2013

Ohio's Students Aren't Showing Much Progress, According to National Test Results
ideastream—No Author Listed
Published November 7, 2013, 6:06 p.m.

Indiana sees rise in test scores on 'nation's report card'
News and Tribune—Maureen Hayden
Published November 7, 2013

Idaho eighth-graders beat national average in reading proficiency
Idaho Statesman—Bill Roberts
Published November 7, 2013

NJ Students Score Well in 'Nation's Report Card' [AUDIO]
New Jersey 101.5—Dino Flammia
Published November 8, 2013, 5:48 a.m.

Dayton Discussing Student Test Scores
KNSI (Minn.)—No Author Listed
Published November 7, 2013, 4:53 a.m.

Report: Wash. Students Among Top 5 in U.S. Math, Reading
KPIU—Florangela Davis
Published November 7, 2013, 7:01 a.m.

Report: SD 4th graders below average in reading
KSFY—Denise DePaolo
Published November 7, 2013, 7:48 p.m.; updated November 14, 2013, 7:54 p.m.

How Texas Students Scored on the Nation's Report Card
KUHF—Laura Isensee
Published November 7, 2013, 10:11 a.m.

State test scores remain steady
KWSN (Pierre, S.D.)—Mark Brown
Published November 8, 2013, 7:02 a.m.

Nevada students score poorly on Nation's Report Card
Las Vegas Review-Journal—Trevon Milliard
Published November 7, 2013, 7 a.m.; updated 5:28 p.m.

Nevada students register gains on 'nation's report card,' but results remain well below average
Las Vegas Sun—Paul Takahashi
Published November 7, 2013, 7 a.m.

Latinos Show Modest Gains in 'National Report Card,' Performance Gap Remains
The Latino Post—No Author Listed
Published November 13, 2013, 2:09 p.m.

Kansas students score high on 2013 reading, math tests
Basehor Sentinel—Peter Hancock
Published November 7, 2013, 11:12 a.m.

Georgia's math, reading scores grow for 2013; Alabama far below national average
Ledger-Enquirer (Columbus, Ga.)—Adam Carlson
Published November 7, 2013

Kentucky fourth-, eighth-graders 'holding steady' on Nation's Report Card
Lexington Herald-Leader—Jim Warren
Published November 7, 2013

California student scores in reading, math climb, but news isn't all good
Los Angeles Daily News—Barbara Jones
Published November 7, 2013, 9:11 a.m.

Education official: Hawaii proved 'lot of skeptics wrong'
The Maui News—Jennifer Sinco Kelleher
Published November 8, 2013

'Nation's Report Card' shows some progress in reading, math
McClatchyDC—Renee Schoof
Published November 7, 2013

NY students not improving in reading, math
Metro.us (New York edition)—Laura Shin
Published November 8, 2013

Black students near bottom in nation on benchmark math, reading test
Journal Sentinel (Milwaukee)—Lydia Mulvany
Published November 8, 2013

Education commissioner points to achievement-gap progress
Minnesota Public Radio—Tom Weber
Published November 7, 2013, 10:48 a.m.

Achievement gap closing some, says Minnesota education commissioner
MinnPost—Brian Lambert
Published November 7, 2013

Tennessee Gets Most Improved In Education, Still Far From U.S. Leader
WPLN (Nashville Public Radio)—Blake Farmer
Published November 7, 2013

Is State Still Pedal to the Metal on Education Reform?
Nashville Scene—Andrea Zelinski
Published November 8, 2013, 8 a.m.

Nation's Report Card shows slight Hispanic gains
NBC Latino—Suzanne Gamboa
Published November 7, 2013, 12:15 p.m.

'Nation's Report Card:' N.H. Students Rank High In Math, Reading Proficiency
New Hampshire Public Radio—Michael Brindley
Published November 7, 2013, 12:45 p.m.

Louisiana ranks low on nation's report card
The Advocate (Baton Rouge, La.)—Will Sentell
Published November 8, 2013

Report: La. ranks low in math, reading
The Advocate (Baton Rouge, La.)—Will Sentell
Published November 8, 2013

National reading, math scores shoot up
Newsday (Long Island)—John Hildebrand
Published November 7, 2013, 10:06 p.m.

NAEP shows Florida fourth-graders perform well in reading
News-Press—Ashley Smith
Published November 7, 2013

New York kids' statewide reading, math scores are middling
New York Daily News—Ben Chapman
Published November 8, 2013, 12:46 a.m.

60% of NY students not up to par
New York Post—Andy Soltis
Published November 7, 2013, 9:59 p.m.

State student performance on national tests remain unchanged
The News & Observer—Lynn Bonner
Published November 7, 2013

Scores On Respected National Exam Say NY Students Not Making Much Progress
NY1—Lindsey Christ
Published November 7, 2013, 9:52 p.m.

Nebraska, Iowa students among best on national tests, but there's room for improvement
Omaha World-Herald—Joe Dejka
Published November 7, 2013, 9:10 a.m.; updated 2:32 p.m.

CA students score among lowest in nation
Orange County Register—Elysse James
Published November 7, 2013, 6:40 p.m.

Oregon students' reading and math skill stuck mostly at average levels, national exam results say
The Oregonian—Betsy Hammond
Published November 7, 2013, 7:03 a.m.; updated 9:31 p.m.

Florida students improve on math, reading skills
Orlando Sentinel—Leslie Postal
Published November 7, 2013, 10:39 a.m.

Texas public school students' reading skills still lag behind national average
Pegasus News (Medill News Service)—Bryan Lowry
Published November 8, 2013

Eighth-grade reading scores improve
The Pine Bluff Commercial—John Lyon
Published November 7, 2013, 8:49 p.m.

Pa. 4th, 8th graders score higher than national average on math, reading exams
Pittsburgh Tribune-Review—Megan Harris
Published November 7, 2013, 10:09 a.m.

Maine students score high on national reading, math tests
Portland Press Herald—Noel K. Gallagher
Published November 8, 2013

South Carolina doesn't see significant increases on national exam while nation makes progress
The Post and Courier—Diette Courrege Casey
Published November 8, 2013, 12:01 a.m.

New York students make small gains on 'the nation's report card,' but most remain below standards
The Post-Standard—Paul Riede
Published November 7, 2013, 2:32 p.m.

RI students score above the national average on national test
Providence Journal—Linda Borg
Published November 7, 2013, 10:01 a.m.

R.I. students show improvement in national math, reading tests
Providence Journal—Linda Borg
Published November 7, 2013, 11:30 p.m.

Kostrzewa - RI test scores better, still not good enough to help economy
Providence Journal—John Kostrzewa
Published November 8, 2013, 3:59 p.m.

National test shows small improvement in state math and reading scores
Radio Iowa—Dar Danielson
Published November 7, 2013

N.J. students again rank near top on 'nation's report card'
The Record—Leslie Brody
Published November 7, 2013, 1:15 p.m.

RI Sees Little Improvement on National Testing
Rhode Island Public Radio—Elisabeth Harrison
Published November 7, 2013, 10 a.m.

Report card: Utah's minority students making gains in reading
The Salt Lake Tribune—Kristen Moulton
Published November 7, 2013, 9:22 a.m.; updated 9:59 p.m.

California students score at bottom of nation in reading, math
San Jose Mercury News—Sharon Noguchi
Published November 7, 2013 10:03 a.m.

California Students Achieve Higher Test Scores On Nation's Report Card
Santa Clarita News—Jeanina Joseph
Published November 7, 2013, 12:57 p.m.

'Nation's Report Card' Shows Students' Math, Reading Skills Slowly Improving
School Library Journal—Karyn M. Peterson
Published November 7, 2013

Schools Chief Tom Torlakson Announces California Students Make Major Gains in 2013 National Tests
Sierra Sun Times—No Author Listed
Published November 8, 2013

Minnesota school achievement gap narrows for some, test shows
Pioneer Press (St. Paul, Minn.)—Christopher Magan
Published November 7, 2013, 12:01 a.m.

Hawaii public school test scores above average
Honolulu Star-Advertiser—Nanea Kalani
Published November 7, 2013, 1:44 p.m.

New results from 'Nation's Report Card' show slight improvement in reading and math
The Star-Ledger—Peggy McGlone
Published November 8, 2013, 6:30 a.m.; updated 6:32 a.m.

Minnesota 4th-graders tie for the top in national math tests
Star Tribune—Kim McGuire
Published November 7, 2013, 10:02 p.m.

Charting Florida's Progress on 'The Nation's Report Card'
StateImpact Florida—John O'Connor
Published November 7, 2013, 2:34 p.m.

How Indiana Students Fared On The Tests the Whole Country Cares About
StateImpact Indiana—Kyle Stokes
Published November 7, 2013, 10:01 a.m.

Florida students achieve small gains in national test
Tampa Bay Times—Jeff Solochek
Published November 7, 2013, 10:02 a.m.

Report: Tennessee shows greatest math, reading gains in nation
The Tennessean—Joey Garrison
Published November 7, 2013, 12:20 p.m.

Tennessee students show big gains on national exam
Times Free Press (Chattanooga, Tenn.)—Kevin Hardy
Published November 7, 2013

Reading, math scores climb in state, but still lag behind national average
Tulsa World (Okla.)—Kim Archer
Published November 7, 2013, 3:25 p.m.

D.C. Students Outpace Peers In Math, Reading Gains
WAMU—Kavitha Cardoza
Published November 7, 2013

Gov. Haslam lauds school improvements, says teachers are satisfied
WATE (Knoxville, Tenn.)—Gene Patterson
Published November 8, 2013, 5:39 p.m.

NAEP tests show Md. student scores still above average
WBAL—Katie Lange
Published November 7, 2013, 6:01 p.m.

TN is fastest improving state in reading and math
WBIR—WBIR Staff
Published November 7, 2013, 1:40 p.m.

Maryland Students Score Above Average
WBOC—Chris Messick
Published November 7, 2013, 7:25 p.m.

New standards bring lower test scores statewide
WECT (Wilmington, N.C.)—Kaitlin Stansell
Published November 7, 2013, 10:07 p.m.; updated November 11, 2013, 10:09 p.m.

School test scores released, show results of new curriculum
WCBD (Chapel Hill, N.C.)—Justin Quesinberry
Published November 7, 2013, 11:09 a.m.; updated 2:08 p.m.

Ohio Students Show Slight Improvement On National Test
WCBE—Jim Letizia and AP and Ohio Public Radio
Published November 8, 2013, 6:47 a.m.

PA Students Show Slight Progress in New Report Card
WESA—Kevin Gavin
Published November 7, 2013, 5:04 p.m.

Bennett Doesn't Take Credit, But Says Reforms Lead To Test Score Gains
WIBC (Ind.)—Ray Steele
Published November 7, 2013

Kansas test scores unchanged
Voice for Liberty in Wichita—Bob Weeks
Published November 7, 2013

TN leads nation in ed. gains
The Wilson Post—Sabrina Garrett
Published November 8, 2013

Education Expert Says Wisconsin Needs Wider Education Policy Discussion
Wisconsin Public Radio—Amanda Magnus
Published November 7, 2013, 3:06 p.m.

NY students show signs of improvement
WIVB—Colleen Hannon
Published November 8, 2013, 7:06 a.m.

D.C. reading and math scores make major gains versus national average
WJLA—Sam Ford
Published November 7, 2013, 11:05 a.m.

Report shows WI achievement gap worst in nation
WKOW—No Author Listed
Published November 8, 2013, 7:04 a.m.

New York State needs work on reading and math
WNYT—WNYT Staff
Published November 7, 2013, 5:28 p.m.; updated 8:35 p.m.

NAEP Test Score: NJ Near Top in Nation!
Woodbridge Patch—Tom Maras
Published November 8, 2013, 7:53 a.m.

Kentucky students scores on par for reading, slightly below for math
WPSD Local—Amanda Roberts
Published November 7, 2013, 9:38 a.m.

Tennessee Schools See Historic Improvement
WREG—Adam Hammond
Published November 7, 2013, 4:38 p.m.

State Superintendent responds to 2013 NAEP report
WTVA (Jackson, Miss.)—No Author Listed
Published November 7, 2013, 9:52 a.m.; updated 10:15 a.m.

DC schools make the grade in test scores
WUSA—Surae Chinn
Published November 7, 2013, 5:29 p.m.

"Nation's Report Card:" Wisconsin has the Widest Achievement Gap
WUWM (Milwaukee)—Ann-Elise Henzl
Published November 7, 2013, 4:40 p.m.

State Officials Hope to Solve Achievement Gap Problem
WUWM (Milwaukee)—Marti Mikkelson
Published November 7, 2013, 1 a.m.

Virginia 4th graders among nation's best readers
WVVA—Greg Carter
Published November 7, 2013, 4:14 p.m.

New York Students Show Improvement
WNYF—No Author Listed
Published November 7, 2013, 11:33 a.m.; updated 11:38 a.m.

2013 Nation's Report Card: Wyo. has "room to improve"
Wyoming Tribune-Eagle—Aerin Curtis
Published November 7, 2013

Local Groups Work To Improve Columbus Reading Scores In Face Of Latest Report
WBNS (Ohio)—Staff
Published November 8, 2013, 4:50 p.m.; updated 5:53 p.m.

Gov. Haslam: TN is fastest-improving state in education, according to NAEP results
WSMV—Josh DeVine
Published November 7, 2013, 5:49 p.m.

NATIONAL ASSESSMENT GOVERNING BOARD
RELEASE PLAN FOR THE
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

The Nation's Report Card: Trial Urban District Assessment in Mathematics and Reading 2013

The 2013 NAEP Trial Urban District Assessment (TUDA) Mathematics and Reading Report Cards will be released together to the general public during December 2013 in one event, as approved by the Board at the December 2013 meeting. Following a review and approval of the report's results, the release will be arranged as an online webinar. The release event will include a data presentation by the Commissioner of Education Statistics, with moderation and comments by at least one member of the National Assessment Governing Board and an additional panelist with expertise in education and assessment matters in large city school districts. Full accompanying data will be posted on the Internet at the scheduled time of release.

The 2013 NAEP TUDA Report Cards in mathematics and reading will present findings from a representative sample of about 365,000 4th-grade and 330,000 8th-grade public school students in 21 urban districts: Albuquerque Public Schools, Atlanta Public Schools, Austin Independent School District, Baltimore City Public Schools, Boston School District, Charlotte-Mecklenburg Schools, Chicago Public Schools, Cleveland Metropolitan School District, Dallas Independent School District, Detroit Public Schools, District of Columbia Public Schools, Fresno Unified School District, Hillsborough County (FL) Public Schools, Houston Independent School District, Jefferson County Public Schools (Louisville, KY), Los Angeles Unified School District, Miami-Dade County Public Schools, Milwaukee Public Schools, New York City Department of Education, San Diego Unified School District, and School District of Philadelphia.

Results will be compared to those of the nation and to a large-city average that includes public schools located in the urbanized areas of cities with populations of 250,000 or more. Data will be presented for all students by such factors as race/ethnicity, achievement gaps, and eligibility for the National School Lunch Program. Contextual information (i.e., student, teacher, and school survey data) with findings of interest will also be reported. Main findings will be included in a brief report summary, with the majority of trends and findings posted in charts and graphs on the new Nation's Report Card website.

DATE AND LOCATION

The release event for the media and the public will occur in December 2013. The release date will be determined by the Chair of the Reporting and Dissemination Committee, in accordance with Governing Board policy, following acceptance of the final report.

EVENT FORMAT

- Introductions and opening statement by a National Assessment Governing Board representative
- Data presentation by the Commissioner of Education Statistics
- Comments by at least one Governing Board member
- Comments by at least one expert in the field of education and assessment matters in large-city school districts
- Questions from the webinar audience
- Program will last approximately 75 minutes
- Event will be broadcast live over the Internet, and viewers will be able to submit questions electronically for panelists. An archived version of the webinar, with closed captioning, will be posted on the Governing Board website at www.nagb.org along with other materials such as the press release and panelist statements.

EMBARGOED ACTIVITIES BEFORE RELEASE

In the days preceding the release, the Governing Board and NCES will offer access to embargoed data via a special website to approved U.S. Congressional staff in Washington, DC; approved senior representatives of the National Governors Association and the Council of Chief State School Officers; and appropriate media as defined by the Governing Board's Embargo Policy. A conference call for journalists who signed embargo agreements will be held to give a brief overview of findings and data and to answer questions from the media.

REPORT RELEASE

The Commissioner of Education Statistics will publicly release the report at the NAEP website—http://nationsreportcard.gov—at the scheduled time of the release event. An online copy of the report, along with data tools, questions, and other resources, will also be available at the time of release on the NAEP site. An interactive version of the release with panelists' statements, a Governing Board press release, subject frameworks, and related materials will be posted on the Board's web site at www.nagb.org. The site will also feature links to social networking sites and audio and/or video material related to the event.

ACTIVITIES AFTER THE RELEASE

The Governing Board's communications contractor, Reingold, will work with Board staff to coordinate a post-event communications effort to extend the life of the results and provide value and relevance to stakeholders with an interest in student achievement and assessment in the nation's large, urban school districts.

Introductory Note: Contextual Information Framework

The revisions proposed to the Background Information Framework for NAEP are intended to make it conform to the Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting, which the Board adopted in August 2012. The changes were recommended by the Ad Hoc Committee on NAEP Background Information that completed its work in August 2013. The Ad Hoc Committee proposed that the term contextual be used rather than background to avoid any misunderstanding that questionnaires were improperly intrusive. The original framework was adopted in 2003. The Ad Hoc Committee felt that although the framework needs updating, its approach and most details remain sound.

Also attached is an annotated version of the August 2012 resolution, showing the pages in the framework where changes are proposed.

A new foreword explains the key changes.

Contextual Information Framework for the National Assessment of Educational Progress

National Assessment Governing Board
Adopted August 1, 2003
Revised December 2013 -DRAFT

National Assessment Governing Board

Darvin M. Winick
Chair

John H. Stevens
Chair, Ad Hoc Committee on Background Questions

Charles E. Smith
Executive Director

Lawrence Feinberg
Project Officer

Background Framework Project

Mary Lyn Bourque, Writer/Editor
Mid-Atlantic Psychometric Services

David Grissmer
RAND

Consultants

Paul E. Barton
Ina V.S. Mullis
Michael Podgursky
Mark Reckase
Herbert J. Walberg

Table of Contents

ADD: Foreword – 2013

Formatted: Font: 12 pt

Foreword – December 2013

In October 2011, eight years after adoption of the NAEP Background Information Framework, the National Assessment Governing Board convened an expert panel to study the NAEP contextual questions and recommend possible changes. The six-member group was chaired by Marshall S. Smith, former dean of the Graduate School of Education at Stanford University and a former U.S. Under Secretary of Education. The panel's report, presented to the Board in March 2012, called the contextual questions "a potentially important but largely underused national resource." (Smith, et al [2012]. *NAEP Background Questions: An Underused National Resource. A Report to the National Assessment Governing Board by the Expert Panel on Strengthening the NAEP Background Questions)*

The report described the information gathered through background questionnaires as "a rich collection of student, teacher and school responses…that can help in understanding the context for NAEP achievement results and give insights into how to improve them." But it said over the past decade the questionnaires had been cut back and little used in NAEP reports. It urged NAEP to "restore and improve upon" its practice of the early 1990s by "making much greater use of contextual data, but do so in a more sound and research-supported way."

With "proper attention," the expert panel declared, NAEP's contextual data "could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education."

After gathering public comment, the Governing Board adopted a Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting. The policy, approved in August 2012, was based on recommendations by the expert panel and provided for an important change in emphasis:

- NAEP reporting should make greater use of contextual data in both regular Report Cards and special focused reports.

- The reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Such information will enrich NAEP reporting, but care should be taken not to suggest causation.

- Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions and their connection to student achievement.

- Modules will be prepared for special studies to provide descriptive information on issues of current policy interest.

- NAEP will include contextual questions from international assessments to obtain direct comparisons of states and TUDA districts with educational practices in other countries.

The Board resolution included a set of implementation guidelines. It also established an ad hoc committee, which reviewed the framework. The committee felt the approach adopted in 2003 and most of the details remain sound, but recommended some updating. The revisions are based largely on the resolution and are incorporated in the text that follows. As NAEP makes the transition from paper-and-pencil to a computer-delivered assessment, the Board hopes the new technology will help make possible the range of topics and flexibility in sampling envisioned a decade ago while limiting the burden on students and schools.

Note on Terminology

The document that follows has been renamed the NAEP Contextual Information Framework. The change was made, from background information framework—the title used in 2003—to avoid any misunderstanding that the information provided and the questionnaires from which it derives are overly intrusive or constitute a "background investigation." Both the statute authorizing the National Assessment and the policies of the National Assessment Governing Board make it clear that this must not be the case.

By law, NAEP is authorized only to collect contextual information that is "directly related to the appraisal of academic achievement, and to the fair and accurate presentation" of assessment results. NAEP must not evaluate or assess "personal or family beliefs and attitudes." The assessment may not disclose "personally identifiable information" and cannot report data on individual students or schools. Under Board policy, adopted in 2002 and retained in the 2003 framework and new update, any questions on student attitudes toward school or various academic subjects, such as reading or science, must be "non-intrusive and have a demonstrated relationship to academic achievement."

In the text of the updated framework the terms contextual and background are used interchangeably, though contextual is the most common and preferred terminology.

Preface

by the National Assessment Governing Board

The National Assessment of Educational Progress (NAEP) has been established by law to monitor the academic achievement of American students. In addition to its academic assessments, NAEP has collected information from hundreds of non-cognitive or contextual questions about students, their educational experiences in class and at home, their teachers, and their schools. Some of these questions provide data for NAEP's reporting categories, but far more have been used to give context to NAEP results or to track factors associated with academic achievement. Some have been used by scholars in social science research.

Concerns have been raised about the selection of contextual variables, the quality of the information obtained, and the validity of inferences drawn from it. There is also concern about the burden that collecting contextual information places on respondents and on the NAEP program. After the National Assessment Governing Board was granted final authority over the background questions in early 2002, it adopted a policy to focus NAEP contextual data on the primary purpose of the National Assessment—to provide sound, timely information on the *academic* achievement of American students. The Board also initiated a process to prepare a general framework to guide the collection and reporting of contextual data.

It is important to understand the National Assessment is not designed to prove cause-and-effect relationships; it cannot prescribe what should be done. But its descriptions of the educational circumstances of students at various achievement levels—considered in light of research from other sources—may provide important information for public discussion and policy action. Used with other research, the contextual data collected by NAEP may give insights into how achievement can be improved as well report to the public on how school personnel and resources related to achievement are distributed.

This framework defines the purpose and scope of NAEP's system of collecting contextual information, including background questionnaires and other sources of non-cognitive data. It establishes criteria for reporting contextual information as part of the National Assessment. The approach it suggests provides for asking various groups of questions to various samples of students at various times.

The framework reflects the following key principles:

- The selection of contextual topics and questions shall be designed to fulfill all legal requirements for the National Assessment and to carry out decisions regarding what NAEP will report and how to report it.

- Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be "directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information."

- The collection of contextual data shall be designed to obtain information that is objective, valid, reliable, and of consistently high quality.

- The system of contextual data collection shall be efficient and designed to minimize the burden on respondents and on the NAEP program. As much data as possible should be obtained from school records and other reliable data sources.

- These principles shall apply both to the collection of general contextual information and to subject-specific background questions. The frameworks for the latter must be focused and prioritized, indicating a core set of variables for regular reporting and a more comprehensive set to be collected and reported less frequently.

- The priority order for contextual information is as follows: (1) reporting categories, as required by law; (2) contextual factors with a well-established relationship to achievement; and (3) subject-specific information.

There is one other consideration—the ~~new~~ role of the National Assessment in the No Child Left Behind Act of 2001. Under this law, all states receiving federal Title I aid are required to participate every two years in NAEP's state-level samples of reading and mathematics in grades 4 and 8. The results ~~will~~ provide an independent yardstick to compare trends on NAEP with performance on each state's own set of required exams.

Because No Child Left Behind places particular emphasis on closing the persistent performance gaps between various student groups, NAEP must be able to report on changes in achievement for all groups specified by law. Through its contextual questions, the National Assessment ~~might~~ can also provide useful information about the students left behind and those who are ahead of them, including the sorts of schools that high-achieving and low-achieving students attend, the courses they take, the patterns of how they are taught, and the qualifications of their teachers. Over time, such descriptive information will allow NAEP to track changes in contextual and instructional factors related to student achievement and in the distribution of important educational resources.

In sum, the purpose of this Contextual Information Framework is to focus the collection and reporting of background data by the National Assessment and to establish clear priorities and limits. We hope to make it possible that with far fewer non-cognitive questions than it has had in the recent past, NAEP will serve the purposes of law and provide the American public and decision makers with useful information. We are committed to improving the quality of data collected and the reporting of results.

Executive Summary

The National Assessment of Educational Progress (NAEP) is a federally authorized survey of student achievement at grades 4, 8, and 12 in various subject areas, such as mathematics, reading, writing, science, U.S. history, the arts, and foreign languages. The No Child Left Behind Act of 2001 (P.L. 107-110) requires the assessment to collect data on specified student groups, including race/ethnicity, gender, socio-economic status, disability, and limited English proficiency. It requires fair and accurate presentation of achievement data and permits the collection of contextual or descriptive information that is related to academic achievement and aids in fair reporting of results. The intent of the law is to provide representative-sample data on student achievement for the nation, the states, and subpopulations of students and to monitor progress over time.

The National Assessment Governing Board (NAGB) sets policy for NAEP and determines the content framework for each assessment. As a result of the No Child Left Behind Act, the Board is responsible for selecting and approving all of NAEP's non-cognitive or contextual questions, as well as the cognitive items over which it has had final authority since 1988. This Contextual Information Framework will guide the development and selection of non-cognitive topics and questions. It will fulfill the purposes of law and provide a clear statement of Board policy.

When NAEP began in 1969-70, its background information was limited to gender, race/ethnicity, and literacy materials at home. During the 1980s the array of non-cognitive questions expanded greatly, both to provide more contextual information and in an effort—never fully realized—to use the assessment for educational research.

This framework will refocus the collection of non-cognitive variables on NAEP's primary mission: providing a fair and accurate measure of student achievement and on achievement trends over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results. NAEP may contribute to research into improving education policy and practice, ~~but I~~its role in this respect is limited, but, used with other research, the contextual data collected by NAEP may give insights into how achievement can be improved as well report to the public on how school personnel and resources related to achievement are distributed.

~~and the framework is not a comprehensive list of possible factors to explore.~~

Since by law NAEP may only collect information that is "directly related to the appraisal of academic achievement," it must concentrate on non-cognitive variables that are known from other research to have such a relationship. The law also specifically prohibits NAEP from asking about personal or family beliefs and attitudes. These points are emphasized in the Governing Board Policy Statement on the Collection and Reporting of Background Data by the National Assessment (adopted May 18, 2002). That policy is incorporated into this framework. The framework also incorporates the Board's more recent Policy Statement on NAEP Background Questions and the Use of

Contextual Data in NAEP Reporting (adopted August 4, 2012). Both policy statements are included in the appendix.

PRIORITIES

The following priorities for collecting and reporting non-cognitive information should be followed in planning background questionnaires, the frequency with which questions are asked, and the samples from which data are collected.

(1) ***Student reporting categories*** that are required by law must be collected as a regular component of all NAEP assessments. These include race, ethnicity, gender, socio-economic status, disability, and limited English proficiency. A core of SES information should be collected in every assessment, such as type of community and poverty status. An expanded set of SES variables may be included periodically or administered to limited samples. Efforts should be made to develop a composite measure or index of SES.

(2) ***Other factors that provide a context for results*** should be sampled periodically, or on a rotating basis, over several NAEP cycles, although a limited set may be asked in every assessment. Contextual factors may include courses taken, student mobility, school safety and discipline, teacher-related factors such as demographics and experience, other factors related to students and schools, and educationally-relevant variables outside school. Modules should be prepared for special studies to provide descriptive information on issues of current policy interest. ~~Although many non-cognitive variables may be of interest, they must be limited to meet the needs of NAEP reporting.~~ In all cases, ~~they~~ non-cognitive variables must be clearly related to academic achievement or to the fair presentation of achievement results.

(3) ***Subject-specific background information*** should be gathered at the same time that achievement in a subject is assessed. This may include relevant course content and requirements, teacher preparation, and other factors related to student achievement. Questions will not be designed to determine effective practices, but to show patterns and trends of factors of interest, based on previous research. Like the contextual information, most of these variables should be sampled periodically, or on a rotating basis, over several administrations of the subject exam, although a limited core set may be repeated every time the assessment is given.

SELECTION CRITERIA

Key criteria for selecting non-cognitive topics and questions are as follows:

- ***Does the current or proposed non-cognitive variable relate to the primary purpose of NAEP and how?*** The primary purpose of NAEP is to report on the academic achievement of students to the American public. It is not to report on the causes of that achievement. Other surveys with longitudinal data are far better suited to examining causality. NAEP's choice of which non-cognitive variables to measure should be guided by how and to what extent the variables selected will support NAEP's primary mission.

- ***Do the current or proposed non-cognitive variables meet professional standards for reliability and validity?*** The NAEP legislation requires that the assessment "use widely accepted professional testing standards (P.L. 107-110, Sec. 411 (b) (5)." This requirement applies equally to non-cognitive and academic variables.

- ***How stable is the non-cognitive variable from period to period?*** If a variable shows little change from year to year, it should be reviewed to determine whether it should be deleted or used on a periodic basis rather than in every assessment.

- ***If new questions are added, have others been deleted in order to limit the burden and expense of NAEP's contextual questionnaires?*** There will always be pressure to collect more information. Mechanisms must be developed to make sure the burden of background questionnaires does not expand over time.

- ***Does a question address specific behavior rather than conclusions?*** Even for such questions, however, caution is advisable because self-reports are often unreliable.

- ***Will the topic or question meet the test of broad public acceptability and not be viewed as intrusive or prying?*** NAEP's non-cognitive questions are not kept secure, and all of them are to be posted on the Internet. Possible objections should be considered in deciding whether or not a question will be asked.

- ***Does the topic or question deal with a factor in which trends over time are important?***

- ***Will the information obtained be of value in understanding academic performance and taking steps to improve it?*** This is a fundamental issue to be addressed in evaluating all background questions proposed for NAEP.

Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest. NAEP should include

contextual questions from international assessments, such as PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study), to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

DATA COLLECTION

Whenever possible, NAEP should use information from school records and other reliable data collections in order to improve the validity of the information collected and limit the ~~background~~ contextual questionnaires in NAEP itself. In exploring the utility of different data sources, the following criteria should be considered: (1) reliability, (2) universality, (3) currency, (4) respondent burden, (5) logistics, (6) efficiency and cost-effectiveness, and (7) the impact on timeliness of NAEP reporting.

Of the student reporting categories in Priority 1, information on gender, race/ethnicity, disability status, and limited English proficiency shall be collected in a uniform manner in all NAEP samples. NAEP is also required to collect information on socio-economic status. This will continue to be done in all samples, although there may be some variation in the number of factors on which data are obtained with a uniform core and more extensive data gathering in some cases.

Because socio-economic status cannot be measured simply or directly, NAEP has used "proxy" variables, such as eligibility for free or reduced-price lunch (a measure of poverty), parent education, and the number of reading materials in the home. The framework provides that NAEP explore development of a composite index for SES derived from ~~the proxy variables~~information ~~currently~~ collected from students and schools. To the extent that the index can be sharpened by additional data from readily available sources, such as zip codes and the census, this option should also be considered. Occasionally and in limited samples, more extensive SES questions may be asked. Although NAEP may never be able to produce a full composite of SES, based on family income, education, and occupation, efforts should be accelerated to develop and use improved measures of socio-economic status, including an SES index. ~~should be made to find an approximation that is more informative than the current set of proxy variables.~~

For the past two decades, NAEP has collected information on a lengthy list of student, teacher, school, and beyond-school factors that may provide a context for achievement results and are of interest to policymakers, researchers, and the public. Yet, NAEP's design as a cross-sectional survey places serious limitations on the inferences that can properly be drawn from this information. We propose a careful review of the contextual factors in NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. Different questions should be cycled in and out of the assessment periodically, and the use of data from non-NAEP sources should increase. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.

The collection of subject-specific contextual information should be focused, limited, and prioritized as part of the subject-matter frameworks adopted by the Board. For subjects tested regularly at two-year or four-year intervals~~each subject~~ there should be a small core set of background items administered to the full sample each time a subject is assessed. An additional, more comprehensive set of questions should be administered periodically or to smaller subsamples.

Whenever feasible, student assessment samples should be divided (spiral sampling) and contextual questions rotated in different years in order to cover more topics without increasing respondent burden. These practices should be initiated in the assessments of reading and mathematics, which are conducted every two years, and considered for other subject areas if the frequency of testing permits.

Clusters of questions should be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning. These clusters could be administered regularly or rotated across assessment cycles and may be used to construct indexes on topics of interest rather than reporting individual items alone.

Thorough reviews should be regularly conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions should be dropped.

Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in contextual questionnaires and their connection to student achievement. Such frameworks should be updated for each assessment cycle and provide the basis for new topics and questions.

Formatted: No bullets or numbering

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. As much data as possible should be obtained from school records and other reliable data sources. The average individual response time to answer contextual questionnaires for each assessment, as calculated in accordance with Office of Management and Budget (OMB) procedures, shall be limited as follows: ~~ten~~10 minutes for each student on paper-and-pencil tests, 15 minutes per student on computer-based assessments, 20 minutes for each teacher, and 30 minutes for each school. Consideration should be given to increasing student response time on paper-and-pencil questionnaires if deemed practical and productive.

REPORTING

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on patterns and trends of non-cognitive variables known to have a relationship to academic achievement and may contain disaggregated data on school conditions and practices for

various groups of students. Data on courses taken before NAEP assessments (either from transcripts or questionnaires) is of great public interest and can be related to academic results.

In addition, ~~supplemental~~ special reports ~~may~~ should be prepared that focus on particular ~~aspects of the background data collected~~topics of public interest and importance. These reports should feature significant contextual information as well as cognitive results. ~~In all cases, NAEP reports published by the National Center for Education Statistics must not state conclusions as to cause and effect relationships and avoid simplistic presentations that imply best practice.~~

All ~~background~~contextual questions and data collected by NAEP should be posted on the Internet so the public may be able to consider them in discussing results. Complete data files should be made available to researchers for further analysis. In all cases, NAEP reports ~~published by the National Center for Education Statistics~~ must not state conclusions as to cause and effect relationships. ~~and avoid simplistic presentations that imply best practice.~~

RESEARCH

As a cross-sectional survey without longitudinal data, the National Assessment is able to document school conditions and practices. It can report on achievement results. But it cannot properly be used to establish direct cause-and-effect relationships. Still, over the past three decades, NAEP has been part of two important research endeavors—exploring changes in the black-white test score gap since 1970 and seeking to establish the impact of state-level reforms during the 1990s. By monitoring achievement well, NAEP has provided sound data for researchers to use. ~~-~~NAEP results have been critical in identifying research hypotheses. Its contextual variables have added valuable information. Its large data sets have been combined with other information to tease out meaning and policy implications, though NAEP's own reports have properly steered clear of these activities.

The Governing Board believes that by doing its main task of monitoring educational achievement well NAEP can make a valuable contribution to education research. ~~The NCES program of secondary analysis grants for researchers to analyze NAEP data should continue. Educational researchers~~ Researchers should be involved, under the auspices of NCES, in developing NAEP contextual questionnaires, validity studies, and other data collection efforts to carry out the provisions of this framework.

The primary purpose of NAEP is to provide fair and accurate information on student achievement. Its primary audience is the American public. The Governing Board believes that in serving its purpose and audience well, NAEP can contribute to educational research. It welcomes the interest and efforts of researchers.

Chapter One: Introduction

The National Assessment of Educational Progress is the only continuous long-term measure of student achievement in the United States in elementary and secondary schools. Its primary purpose is to report to the American public on academic achievement and its change over time.

Nature and Purpose of NAEP

The NAEP survey consists of two major components: academic assessments that measure the achievement of students on a broad range of content, and non-cognitive survey questions that collect descriptive information from students, teachers, and school administrators about demographic characteristics and the educational process. Since 1969 NAEP has measured achievement in most areas of the school curriculum, including mathematics, reading, writing, science, U.S. history, world geography, civics, economics, foreign language, computer science, and the arts. The content of NAEP assessments is determined through a framework development process that articulates the content parameters for each area and recommends subject-specific non-cognitive areas for data collection and reporting.

NAEP's purpose is to report to the public on the status of academic achievement in America. The assessment does not report results for individual students, but only for groups of test-takers having large, representative samples, e.g., students from rural schools, from various ethnic groups, or from participating states, and, on a trial basis, large urban school districts. It must be able to provide data for fair and accurate comparisons between the states and subgroups on which it reports. The contextual data play a crucial role in ensuring the fair comparisons—over time and between student groups—that are at the heart of NAEP's mission and value.

Nature and Purpose of Contextual Data

The most recent NAEP reauthorization (P.L. 107-110) gives the National Assessment Governing Board "final authority" to approve "all cognitive and non-cognitive assessment items." This framework deals with the non-cognitive side of the Board's responsibility, including the items that identify students in NAEP's required reporting categories and the other information that provides a context for results and tracks factors associated with academic achievement.

The term "non-cognitive," as used in the law, seems more inclusive than the phrase "background questions" by which the collection of non-academic information has been termed by NAEP in the past. However, non-cognitive is also less readily understandable than background or contextual information. In this document the terms will be used interchangeably to refer to all of the information beyond the academic assessment that NAEP uses to make its academic results more meaningful to the public.

When NAEP began, the collection of non-cognitive data was limited to the demographic categories of gender and race/ethnicity, and to two measures of home environment or socio-economic status—level of parents' education and literacy materials in the home. In addition, an index was constructed, based on data from the U.S. Census and a brief school questionnaire, to report achievement results for schools in three types of communities—disadvantaged urban, advantaged urban, and rural.

During the 1980s the use of non-cognitive questions was greatly expanded to accommodate several functions within NAEP (Reckase, 2002). First, they were used to define a more extensive array of subgroups of the student population for reporting purposes. For example, NAEP results are now reported by gender, race/ethnicity, parents' highest level of education, type of school, participation in Title I, and eligibility for free/reduced-price lunch

A second reason for collecting non-cognitive information is to inform educational policy by describing the contexts for learning, sometimes called ~~opportunities~~ opportunity-to--learn (Mullis, 2002). Broadly, this involves the content specified in the curriculum, whether and how that content actually is taught, students' propensity to learn, as well as home and school factors that can enhance learning.

In conjunction with the descriptions of students, contextual information about educational settings and experiences can reveal striking differences in how important aspects of education and educational resources are distributed among different groups. For example, do disadvantaged minority students have less access to science laboratory equipment than more advantaged groups? Do girls take less rigorous mathematics courses than boys? The data on course taking has been used widely to discuss the patterns and trends in mathematics achievement. Having this information as part of NAEP has added to the public impact of assessment results.

A third function of the non-cognitive questions has been to support research into factors that may be related to student achievement. The questions serving this function have sought information not only on curriculum, teaching methods, and discipline in the school, but also on educational activities at home. For example, *The 1998 NAEP Reading Report Card* (Donahue, Voelkl, Campbell, & Mazzeo, 1999) reports on television viewing, daily reading habits, classroom reading and writing assignments, and discussion of schoolwork at home. While secondary researchers have used NAEP to investigate relationships to student achievement, the basic design of the assessment as a cross-sectional survey without longitudinal data limits its usefulness. Research has been most productive when NAEP is combined with other data sources and in descriptive studies that track changes over time.

Non-cognitive data are also necessary to support certain technical functions of NAEP. For example, some non-cognitive information is used to evaluate the potential for bias resulting from non-participation. That is, did the students absent or refusing to participate in the assessment differ in such significant ways from those who did take part that results were changed? Non-cognitive variables also play an important role in

NAEP's sampling and weighting procedures, and sometimes in checking the validity of results. Many of these variables are taken from other data sources, such as the Common Core of Data (CCD), but some come from the administration roster collected from schools prior to testing, the records kept by test administrators, and student questionnaires.

Finally, NAEP non-cognitive questions have been used in the technical process for preparing estimates of student proficiency distributions on the cognitive component of the assessment. But their role in this process is limited to facilitating data analysis. Only the student responses to cognitive questions are used to determine achievement results. Contextual variables are used to define the groups for which cognitive data are reported.

Once test results for a group are determined, the NAEP analytic process makes use of contextual data available to prepare a second data set—identical in its group scores to the first—that can be handled by much simpler computer programs to prepare other analyses and reports. However, only the contextual factors to be reported on are needed for this analytical work, called conditioning. The precision of NAEP results is not reduced if contextual items not used for reporting are eliminated.

This contextual information framework will focus the collection of non-cognitive information on NAEP's primary mission: providing, as the law stipulates, "a fair and accurate measurement of student academic achievement and reporting trends in such achievement" over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results.

Development of NAEP Contextual Information Framework

In the Policy Statement on Redesigning the National Assessment of Educational Progress (adopted in August 1996), the Governing Board sought to improve the validity of contextual information on NAEP, increase the efficiency with which it is collected, and reduce the number of contextual questions in the assessment itself. The statement was based on the report of a Design/Feasibility Team (Forsyth et al, 1996), headed by Robert Forsyth, which recommended a design that would rotate the collection of non-cognitive data into distinct modules administered over several assessment cycles. NAGB endorsed implementing that recommendation through a system of *comprehensive* and *standard* NAEP assessments that would be administered on a cyclical basis (NAGB, 1996).

Standard assessments would ask a short, essential core of contextual questions associated with a content area. Periodically, a *comprehensive* assessment would employ a much fuller complement of such questions to probe that area more extensively. Although some efforts have been made to reduce the contextual questionnaires and streamline data collection, the full impact of the NAGB policy has not yet been realized.

In early 2002, the No Child Left Behind Act transferred final authority over the non-cognitive questions from the National Center for Education Statistics to the National

Assessment Governing Board. The Board adopted a new policy governing the development and selection of non-cognitive questions in May 2002, and initiated a process to prepare a general framework for non-cognitive data (NAGB, 2002). This framework would define the scope of NAEP contextual questionnaires, the priorities for collecting non-cognitive information, and the criteria for reporting non-cognitive data in NAEP. (See Appendix for full text of the policy.)

The Board created an Ad Hoc Committee on Background Questions and conducted an all-day workshop on the NAEP non-cognitive questions on September 24, 2002. Six consultants prepared and presented papers at the meeting that was attended by Board members, academic researchers, representatives of the national teacher organizations and other education groups, and NAEP contractors and staff. The six consultants are identified on the title page as contributors to this document.

In the months after the workshop, a draft framework was prepared. It was refined at several meetings of the Ad Hoc Committee, posted for public comment on the Internet, and was the subject of a public forum in Washington, D.C., on May 1, 2003. Altogether, oral comment and written testimony were received from 22 persons and organizations, many with differing perspectives and views. The Ad Hoc Committee and the Board carefully considered these comments, and the draft framework was revised at a Committee meeting on June 25. The Committee heard additional comment and made final revisions on July 31. The background information framework was reviewed by the full Governing Board several times during the course of its development. The Board adopted it unanimously on August 1, 2003.

While this framework is not a consensus document, it does encompass the thinking of a wide range of researchers, policy analysts, and users of NAEP data. It is the product of discussion and deliberation by the Governing Board, and incorporates Board decisions on the nature and focus of the contextual information to be included in NAEP.

Requirements of NAEP Statute

The No Child Left Behind Act of 2001 (P.L. 107-110) requires NAEP to collect information on gender, race/ethnicity, socio-economic status, disability, and limited English proficiency. It must report test data on these groups, whenever feasible, that is cross-tabulated, compared, and reported according to the categories required.

The law also requires NAEP to collect only information that is directly related to academic achievement and to the presentation of such information in a fair and accurate manner. This means that NAEP needs to concentrate on variables that are known to be related to achievement rather than on theoretical constructs. The statute requires the Governing Board to ensure that all NAEP questions are "free from racial, cultural, gender, or regional bias"—a provision from previous law. But it adds new language that questions must be "secular, neutral, and non-ideological" and must not "evaluate or assess personal or family beliefs and attitudes."

In their report on the bill, the House-Senate conference committee that negotiated its final form says the law "does not preclude the use of non-intrusive, non-cognitive questions, approved by the National Assessment Governing Board, whose direct relationship to academic achievement has been demonstrated and is being studied as part of [NAEP] for the purposes of improving such achievement." The report language is not binding, but is intended to guide implementation of the law. ***This framework emphasizes that the legal prohibitions must be followed in preparing contextual questions and collecting any other non-cognitive data for NAEP***.

In addition, the law makes it clear that NAEP may not disclose any personally identifiable information or maintain any system of records that contains such data. These restrictions are not new. They have dictated careful procedures that must be continued.

Purpose and Rationale of Contextual Information Framework

The purpose of the framework for contextual information is similar to that of NAEP's content area frameworks: to guide the development of the assessment. The content frameworks have described the topics to be tested by NAEP and provided an outline of the assessment for each subject area. Purposefully, the frameworks attempt to be independent of a particular pedagogy. They do not specify what educational resources or processes should be used, but rather describe important achievement results. They provide states, schools, policymakers, and the public with a logical outline of the approach used in constructing the assessment.

The framework for NAEP contextual data will specify the parameters of the assessment from a reporting perspective. The contextual information that NAEP uses in its reports helps to give context and meaning to the cognitive results. It must be collected in a systematic way from the NAEP testing samples either through questionnaires or from other reliable sources, such as school records and other federal surveys. Collecting descriptive information from a variety of sources can improve the quality of the data obtained and increase efficiency while reducing the burden on respondents.

The Governing Board adopted a Policy Statement on the Collection of Reporting of Background Data on May 18, 2002 (NAGB, 2002). The statement is incorporated into this framework and attached in the Appendix. A further statement, entitled Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting, was adopted by the Board on August 4, 2012. It has been used in revising the framework text and has been included in the Appendix.

Chapter Two: Priorities and Criteria For Collecting and Reporting Non-cognitive Data on NAEP

This chapter presents priorities for collecting and reporting non-cognitive information on NAEP. It also includes the criteria for selecting particular topics and questions, and for determining the frequency with which various data elements are reported. A final section presents criteria for identifying and selecting contextual data sources.

Priorities for Non-Cognitive Information

The following priorities for collecting and reporting non-cognitive information are based on legal requirements, the purposes of NAEP, and the strengths and limitations of the assessment. They should be followed in planning contextual questionnaires, the frequency with which questions are asked, and the samples from which data are collected.

(1) ***Student reporting categories that are required by law must be collected as a regular component of all NAEP assessments.*** These include race, ethnicity, gender, socio-economic status, disability, and limited English proficiency. A core of SES information should be collected in every assessment, such as type of community and poverty status. An expanded set of SES variables may be included periodically or administered to limited samples. Efforts should be made to develop a composite measure or index of SES.

(2) ***Other factors that provide a context for results should be sampled periodically, or on a rotating basis~~,~~ over several NAEP cycles, although a limited set may be asked in every assessment.*** Contextual factors may include courses taken and course requirements, student mobility, school safety and discipline, teacher-related factors such as teacher demographics, preparation, credentials, and experience, and other factors related to students, schools, and educationally-relevant variables beyond the school. Modules should be prepared for special studies to provide descriptive information on issues of current policy interest.~~Although these types of non-cognitive variables are of interest, they must be limited so that they meet the needs of NAEP reporting~~. In all cases, ~~they~~ non-cognitive variables must be clearly related to academic achievement or to the fair presentation of achievement results.

(3) ***Subject-specific information may be gathered at the same time that academic achievement in a particular area is assessed***. This may include relevant course content and requirements, teacher preparation, and other factors related to achievement in the subject assessed. Questions will not be designed to determine effective practices, but to show the patterns and trends of factors of interest, based on previous research. Like other contextual information, most of these variables should be sampled periodically, or on a rotating basis, over several administrations of the subject exam, although a limited core set may be repeated every time the assessment is given.

With regard to the points above, Walberg (2002) makes a suggestion that might be a workable solution to consider. Just as students in the NAEP samples do not respond to all the questions, say, in reading, but only to a portion of those for any one grade-level, so too, the non-cognitive questions could be rotated through different (smaller) NAEP samples. These non-cognitive "testlets" could be rotated through the NAEP samples by class or school, with students receiving different, expanded "testlets" in addition to a core set of contextual questions.

Criteria for Selecting Non-cognitive Topics and Questions

The Advisory Council on Education Statistics (ACES), a technical panel that used to advise the National Center for Education Statistics, spent a considerable amount of effort on the issue of NAEP non-cognitive questions. Its guidelines, adopted in May 1997, include a set of key questions that should be utilized in selecting topics and questions for NAEP contextual data collection. The questions with commentary are summarized below:

- ***Does the current or proposed non-cognitive variable relate to the primary purpose of NAEP and how?*** The primary purpose of NAEP is to report on the academic achievement of students to the American public. It is not to report on the causes of that achievement. Other surveys with longitudinal data are far better suited to examining causality. NAEP's choice of which non-cognitive variables to measure should be guided by how and to what extent the variables selected will support NAEP's primary mission.

- ***Do the current or proposed non-cognitive variables meet professional standards for reliability and validity?*** The NAEP legislation requires that the assessment "use widely accepted professional testing standards (P.L.107-110, Sec. 411 (b) (5)." This requirement applies equally to non-cognitive and academic variables. It is already known that some non-cognitive variables in NAEP have weak reliability (e.g., data from 4th graders on their parents' highest level of education and the self-reports of teachers on classroom practice). If more reliable sources of such data cannot be found, these variables should be deleted from the assessment.

- ***How stable is the non-cognitive variable from period to period?*** If a variable shows little change from year to year, it should be reviewed to determine whether it should be deleted or used on a periodic basis rather than in every assessment.

- ***Is the proposed or current non-cognitive variable of timely interest?*** The educational environment changes from time to time, and consequently public interest in particular variables will change as well. It would serve NAEP well to review the set of non-cognitive variables periodically with this criterion in mind, deleting those that do not meet the test of timeliness and substituting others of current interest.

- ***If new questions are added, have others been deleted in order to limit the burden and expense of NAEP's contextual questionnaires?*** There will always be pressure to collect more information. Mechanisms must be developed to make sure the burden of contextual questionnaires does not expand over time.

- ***Does a question address specific behavior rather than conclusions?*** For example, a question that asks teachers whether they adhere to national standards in mathematics or another subject is conclusionary and hard to interpret, since many teachers are apt to say yes, regardless of what they do. It would be better to ask about specific behaviors, such as homework assignments or computer use. Caution is advisable in this area too because self-reports are often unreliable.

The Board believes three other important criteria must also be considered:

- ***Will the topic or question meet the test of broad public acceptability and not be viewed as intrusive or prying?*** NAEP's non-cognitive questions are not kept secure and must readily be available to anyone requesting a copy. Under Board policy, all questions asked are to be posted on the Internet. Possible objections should be considered in deciding whether or not to ask them.

- ***Does the topic or question deal with a factor in which trends over time are of importance?*** If trends are deemed important and the factor is related to achievement, the topic or question should be included periodically on a four-year or eight-year cycle, rather than being part of the contextual questionnaire each year. For example, measuring television watching in every NAEP assessment is not necessary. But it can be valuable to measure TV-watching every four or eight years to find out whether or not it is increasing.

- ***Will the information obtained be of value in understanding academic performance and taking steps to improve it?*** This is a fundamental issue to be addressed in evaluating all contextual questions proposed for NAEP.

Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest. NAEP should include contextual questions from international assessments, such as PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study), to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

Formatted: Indent: Left: -0.06", First line: 0.56"

Criteria for Selecting Data Sources

NAEP has collected non-cognitive information from students, teachers, and schools, using NAEP contextual questionnaires. There are also administration rosters, completed by test administrators at the school level in advance of testing to determine characteristics of the testing samples. The Common Core of Data (CCD) is used to identify characteristics of schools (e.g., Title I funding), and schools also complete a questionnaire on special needs students (e.g., students with disabilities and limited English proficiency).

However, the collection of non-cognitive data may be shifted among these sources or to new sources in order to improve reliability, increase efficiency, or reduce burden. State management information systems and data collected for school report cards, ~~as required by the No Child Left Behind Act, may~~ have become ~~very~~ increasingly useful for NAEP. ***Whenever possible, NAEP should use information from school records and other reliable data collections about students and schools in order to improve the validity of the information collected and limit the contextual questionnaires in NAEP itself.***

In exploring the utility of different data sources, the following criteria should be considered:

- ***Validity*** – Is the data obtained from the new source a valid indicator of what it purports to measure?
- ***Reliability*** – Is the data from the new source at least as reliable and consistent as that from the source previously used?
- ***Universality*** – Can the required data be collected by this method for all (or almost all) of the students and schools participating in NAEP and will it support valid comparisons over time?
- ***Currency*** – Will data obtained from a new data source be current enough to relate clearly to the assessment being conducted? If data from the census or some other source is several years old it may not accurately describe school or neighborhood conditions at the time of testing.

- ***Respondent Burden*** – Will the new source(s) reduce the burden on students, teachers, and schools in filling out NAEP questionnaires? Will the total amount of respondent burden be decreased?

- ***Logistics*** – Will the alternative source(s) be logistically possible, or will there be more logistical problems than with the previous data source? Logistics includes such considerations as cost, time, administrative personnel resources, and steps needed to ensure accurate coding and data analysis.

- ***Efficiency and cost-effectiveness*** – How efficient will the new data source be in comparison to the previous one? For example, it may be more efficient to collect data from a state management information system about the state's schools, teachers, or students, rather than obtaining it from the test samples directly, but efficiency and cost-effectiveness should be determined before a change is made.

- ***Timeliness of NAEP reporting*** – How will a change in data sources affect the speed with which NAEP can be reported? Some changes will speed operations, but those that slow down NAEP reporting are not desirable.

Chapter Three: Topics and Types of Contextual Data

This chapter will cover the non-cognitive topics that are required for reporting under the No Child Left Behind Act of 2001 (P.L. 107-110), as well as those that should be considered for inclusion in NAEP on a cyclical basis. It discusses socioeconomic status (SES), contextual factors of interest to public policy, and subject-specific variables.

Demographic Reporting Categories

The demographic variables ~~currently~~ collected by NAEP come from two sources. Information is obtained from school records on ~~are~~ gender, age, race/ethnicity, and two elements of socio-economic status (SES) — participation in Title I and eligibility for free or reduced-price lunch, which is based on family income. The school~~-~~ records are also used to indicate whether a student is classified as disabled ~~In addition, information is obtained on disability status and on students who are classified as~~or limited English proficient. ~~All of this information is collected on an administration roster, completed from school records in advance of testing~~. In addition, data on race/ethnicity is also collected on the NAEP student questionnaire, and students are asked to report on the highest level of each parent's education and on several aspects of home environment~~-~~, including number of books, internet access, and whether they have their own bedroom.

A more extensive questionnaire is completed by school staff on each student selected for NAEP who is classified as either disabled or limited English proficient (LEP). For students with disabilities (SD), the questionnaire collects data on the specific disability and its severity, the student's Individualized Education Plan (IEP), type of curriculum, whether the student participates in standardized testing (with our without accommodations), and the accommodations allowed on state and district standardized tests in presentation, response, setting, and timing. For LEP students, the questionnaire covers native language, number of years of academic instruction in English, percent of instruction in English and/or native language, and the testing accommodations provided under district or state policy. In the future, NAEP might also identify students who recently exited from LEP programs and track their achievement.

NAEP is required to collect information on all of these categories (except age), but has some discretion in determining definitions and aggregating responses. These data will continue to be collected in a uniform manner in every NAEP assessment, although, for socio-economic status, as explained in the section below, there may be some variation, with a uniform core and more extensive data-gathering in some cases.

Socio-economic Status (SES)

Under current law, NAEP is required to collect information on socio-economic status. SES also is clearly a factor that has been shown to be related to academic achievement in many research studies, beginning with the Equality of Educational Opportunity Commission Report (Coleman et al., 1966). The research community's consensus over the past four decades has been to deal with the influence of SES on other achievement-related variables by holding SES constant while examining the other effects, for example, adjusting for SES while looking at effects of class size or teacher training. NAEP does not adjust for SES, but it does report on the relationship between student achievement and SES proxy variables like parents' education or Title I participation.

NAEP has not been able to measure SES directly, using its present set of questions and data sources, i.e., the student, teacher, and school questionnaires. The assessment has used "proxy variables" for SES, including students' eligibility for the National School Lunch program, participation in Title I, parents' education, and the number of reading materials in the home (newspapers, magazines, books, etc.)—information on the latter two factors being reported by students in the assessment samples. In addition, NAEP uses census data to classify schools into different types of location, based on Census Bureau definitions, such as central city, suburban/large town, and rural/small town. The questions on newspapers and magazines were dropped in the mid-2000s as circulation dwindled, and were replaced by an item on internet access.

Strictly speaking, these are individual proxy variables and are not combined into a composite variable. However, both the questions on parent education and home environment ~~are~~ have been coded in a pseudo-composite manner. For example, the parent education related to the student is the higher of either the mother's or father's education level. On the four home environment questions used until the mid-2000s student responses ~~are~~ were coded differently for a "yes" answer to two questions or fewer, "yes" to three questions, and "yes" to four questions, as well as omitted responses (Allen, Carlson, & Zelenak, 1999).

At the lower grade levels, students' reports of their parents' education are questionable at best, while the National School Lunch program sorts students only into three categories (Yes, No, and Unknown) and Title I into two categories (Yes or No). For many years, NAEP used a reporting category of disadvantaged urban schools, constructed from information provided by school principals. This was discontinued in the mid-1990s because the category lacked a consistent definition from year to year and between different state samples. There also were serious doubts about the reliability of the information on which it was based. The data on eligibility for the National School Lunch Program have also become increasingly problematic because of expansion of the program and administrative changes allowing whole-school or whole-district eligibility in high-poverty areas. In short, there has been considerable concern over many years about the quality of the SES measures in NAEP, both for reporting to the public and for analysis by researchers.

Barton (2002) suggests two alternative approaches for improvement: (1) a composite index for SES, or (2) a parent questionnaire. A composite index is viable

using the same information that is currently collected in NAEP, or perhaps augmented with a few targeted questions or census data, possibly the zip code of student home addresses. ***The necessary analytical work should be initiated through small research studies using extant NAEP data sets in order to check systematically the validity of a composite index as a better measure of SES in NAEP samples. The results could vary by grade level, in which case, adjustments might be needed in the way the data are collected, augmented, and/or confirmed. NAEP may never be able to produce a full composite of income, education, and occupation, but efforts should be accelerated to develop and use improved measures of socio-economic status, including an SES index.*** ~~***made to find an approximation that is more reliable than the current set of individual proxy variables.***~~

In November 2012, an expert panel convened by the National Center for Education Statistics recommended prompt development of an SES composite measure.

Formatted: Font: Not Bold, Not Italic

The argument in favor of this approach is that it advances the goals of the current law without impacting data collection in unforeseen ways. Barton suggests that such an index would enable NAEP to report results in terms of SES quartiles (much the same way that the National Educational Longitudinal Survey, NELS, does). Further, it would allow the assessment to report cross-tabulations on distributions of students in the NAEP achievement level categories by SES. A good measure of SES would improve the monitoring of achievement gaps among various racial/ethnic groups, although sample sizes may not be large enough within all ethnic groups or types of schools. Finally, a composite SES index may be beneficial to states and districts in the Trial District Assessment (TUDA), enabling NAEP to compare the performance of groups of students with the same socio-economic status, which is a factor of high public and policy interest.

The argument against such an approach is that SES would continue to be measured indirectly, i.e., by using proxy variables, albeit through a composite index. There would also be disagreements about precisely which variables to include in the index and how to weight different factors. For example, Armor (D. J. Armor, personal communication, December 18, 2002) has suggested that two variables ~~recently~~ deleted from the NAEP student questionnaire in 2000 be reinstated, namely, the number of siblings in the home and family status (student lives with both parents, mother or father, neither). These variables were dropped because of concerns about intrusiveness, but they may be of considerable importance in constructing an SES index. The item on number of parents in the home was restored in 2013.The Board will have to weigh the considerations involved, and may decide there is value in using them periodically or in limited samples.

A parent questionnaire has been proposed as a more reliable means of collecting SES data than relying on student reports, school records, or census data. Other National Center for Education Statistics surveys, for example, NELS and the Early Childhood Longitudinal Study, have employed parent questionnaires that ask direct questions regarding occupation and income.

However, the National Assessment of Educational Progress involves far more students than any of these research surveys. Accordingly, a parent questionnaire on NAEP would entail far more respondent burden and might arouse more controversy, making it more difficult to accomplish the primary mission of the assessment to measure student achievement. A parent questionnaire has been considered by NAGB in the past, but rejected as too burdensome and intrusive. Because these considerations are still persuasive, particularly as the scope of NAEP has expanded, no work should be undertaken on developing a parent questionnaire.

In sum, because of its importance and the requirements of law, information on socio-economic status must be collected in all NAEP samples, although there may be some variation in the number of factors on which data are obtained. ~~Research~~ Efforts should be ~~conducted~~ made to develop ~~into creating~~ a composite measure or index of SES based on school records and the student questionnaire. To the extent that an index can be sharpened by additional information from readily available sources, such as zip codes and/or census data, this option should be considered as well.

A core of SES information should be collected in every assessment, such as type of community (e.g., central city, rural, etc.), poverty status (e.g., eligibility for free or reduced-price lunch and Title I participation), reading materials in the home, and level of parent education. ~~n though steps~~ Steps must be taken to ensure that such data are reliable. ~~An expanded set of~~Additional SES variables may ~~be included~~also be ~~periodically and~~included, such as number of siblings and parents at home, possession of computers, and parent occupation. Periodically, an expanded set may be administered.

~~***NAEP should explore the use of an SES index derived from proxy variables currently in either the administration roster or student questionnaire. To the extent that an index can be sharpened by additional information from readily available sources, such as zip codes and/or census data, this option should be considered as well.***~~

Formatted: Body Text, Left, Indent: First line: 0.5"

Public Policy Contextual Factors

For the past two decades NAEP has collected information on student, teacher, school, and beyond-school factors that are of interest to policymakers and the public. For students, some of these factors have included course-taking patterns, TV-watching, homework, and use of computers. For teachers, the contextual factors have included educational background, credentials, years of experience, and participation in professional organizations, to name a few.

The lists of factors have been long. They have become burdensome both to respondents and to the efficient scoring, analysis, and reporting of the NAEP survey. The way they have been reported—through simple one-way tabulations—has encouraged unwarranted conclusions about cause-and-effect relationships.

We propose a careful review of the contextual factors on which information is collected by NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. ~~Data should be obtained periodically, on a rotating basis, over several NAEP cycles, although a limited set of factors may be included in every assessment~~Modules should be prepared for special studies to provide descriptive information on issues of current policy interest. ~~Information~~ Data should be collected at meaningful intervals in ways that may show significant patterns and change over time.

Two documents are helpful in surveying the research base and presenting alternatives for NAGB to consider. The first is *Monitoring School Quality: An Indicators Report* (Mayer, Mullens, & Moore, 2001), prepared by Mathematica Policy Research, Inc. for NCES. This report presents a research synthesis, indicating factors for which there is a research base showing a strong relationship to academic achievement. The synthesis, involving a review panel as well as statistical analyses, identifies the following as factors related to student results: the academic skills of teachers, teacher assignments (such as out-of-field teaching), course content, student discipline and school safety, class size, and focus on academic achievement. Other sources of information are available on all of these factors, but only through NAEP can they be related to the achievement of broad groups of students over time.

The second document, *Making Connections* (Greenberg, Stancavage, Farr, & Bohrnstedt, 2001), was prepared for NCES by the American Institutes for Research and presents an elaborate typology of non-cognitive variables that could be measured by NAEP. It is organized into seven broad categories of non-cognitive information related to students, instructional content and practice, teachers, schools, school community factors, beyond school factors, and federal, state, and district policy. The listing goes beyond what NAEP can and should handle, but its discussion is thoughtful and the document is useful for planning.

Subject-Specific Contextual Data

For each subject assessed by NAEP, additional subject-specific contextual information has been collected from students, teachers, and schools. These data fall into the broad category of instructional content and practice. Under that umbrella come such topics as the curriculum taught, course offerings, class management and style, ability grouping, and modes of instruction. Subject-specific data collection has expanded enormously over the past two decades, and in recent years has included five to ten minutes of questions for students, about 30 minutes of questions for teachers, and 30 to 45 minutes for school administrators.

~~Now is the time for t~~These questions ~~to~~ should be focused, limited, and prioritized. Future subject-matter frameworks adopted by the Governing Board should spell out clearly what these priorities will be.

Whenever feasible, student assessment samples should be divided (spiral sampling) and contextual questions rotated in different years in order to cover more topics without increasing respondent burden. These practices should be initiated in the assessments of reading and mathematics, which are conducted every two years, and considered for other subject areas if the frequency of testing permits.

~~The design for doing this was presented to the Board in the 1996 report of a Design/Feasibility Team of prominent researchers (Forsyth, R. et al, 1996). The group recommended that a core set of non-cognitive questions should be administered to students each time a subject is assessed by NAEP. In addition, a more comprehensive questionnaire would be given whenever a new framework is introduced and repeated every eight to ten years. For example, an extensive set of background questions in reading and mathematics (grades 4 and 8) was administered in 2003, the baseline year for the No Child Left Behind legislation. Another complete set should be administered in mathematics in 2005 and in reading in 2007, the years in which revised frameworks are first used, and then should be repeated at an interval of eight years. In the intervening years, only the more limited core modules will be administered.~~ Similar patterns should be ~~established~~ considered for the school and teacher questionnaires.

In ~~The~~ NAEP assessments ~~in other subjects~~ given at intervals of four years or more, such as writing, science, history, ~~geography,~~ and civics, ~~should have~~ a core set of non-cognitive questions should be administered to the full sample, with different sets of longer, more extensive questionnaires being administered to smaller sub samples.

With states ~~now~~ required to participate in NAEP every two years, the total number of students tested has expanded substantially from what it was in the program's first decades~~.~~ This makes even more compelling the case for limiting the NAEP contextual questionnaires and rotating the background questions.

Clusters of questions should be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning. These clusters could be administered regularly or rotated across assessment cycles and may be used to construct indexes on topics of interest rather than relying on stand-alone items only.

~~***NCES should prepare for Board review and approval a plan indicating the frequency, sample size, and schedule of rotation for all background variables and questions on which information is to be collected by NAEP. This should include both questionnaires and alternate data sources to obtain core reporting data, subject-specific information, and data on achievement-related contextual variables from a variety of NAEP samples – national only, national and state, and a subset of the national sample. The plan should indicate the frequency and schedule of rotation for each of the questions proposed. It should also indicate any questions needed for quality control purposes. The recommendations should be prepared with input from researchers and state policy analysts, as appropriate, and updated on a regular basis.***~~

Formatted: Indent: First line: 0.5"

Table 1 presents a model schedule for comprehensive and core sets of subject-related variables through 2013. It is based on the schedule of assessments approved by the Board in May 2003.

Table 1

Model Data Collection Schedule for Comprehensive and Core Sets of Non-Cognitive Variables by Subject Area

Subject Area	Data Collection Year for Comprehensive Set of Variables	Data Collection Year for Core Variables Only
Reading	2003, 2007, 2013	2005, 2009, 2011
Mathematics	2003, 2005, 2013	2007, 2009, 2011
Foreign Language (12)	2004, 2012	
World History (12)	2010	TBD
Economics (12)	2006	TBD
Civics	1998, 2012	2006
Writing	2002, 2011	2007
Arts (8)	1997, 2008	
Science	2000, 2009	2005
US History	2001, 2006	
Geography	2001, 2010	

NOTE: Based on schedule approved by NAGB on May 17, 2003.

31

Chapter Four: Non-cognitive Data Sources and Collection

This chapter discusses the sources of non-cognitive information for NAEP and the reporting categories that the information describes. It includes a NAEP Contextual Information Matrix, organized by priorities, which summarizes the types of descriptive information NAEP collects, reporting units, and data sources.

NAEP Student, Teacher, and School Samples

The NAEP student samples vary in size and purpose. Their overall total has become very large. Starting in 2003, national NAEP samples are specified at the state and jurisdictional levels, with approximately 3,000 students per subject and grade (4 and 8 only) for each of the 50 states, plus the District of Columbia, and Department of Defense domestic and overseas schools. Puerto Rico (in mathematics only) has a sample of about 3,000. In addition, the ten Trial Urban District Assessment (TUDA) districts have sample sizes of the order of 3,000 to 5,000 each. There also are a nationally-representative sample of charter schools, totaling about 3,000 students, and national private school samples totaling about 12,000 in each grade.

At grade four, therefore, the total NAEP sample approximates 436,000 students. The grade eight sample is about the same at 432,000 (excepting charter schools). The grade 12 sample is for a pilot test and includes only about 6,000 students (Rust, 2002). In most future years the twelfth grade samples are expected to have about 30,000-40,000 students assessed in national samples only for three subjects.

In addition to the nearly one million students tested, about 80,000 teachers of those students complete teacher questionnaires and some 13,000 schools complete school questionnaires. Several thousand school districts also supply data for the assessment. The sampling and weighting procedures in NAEP use data from the CCD files as well as census data and school-level achievement data from the states for improving NAEP stratification procedures. The NAEP non-cognitive data collection effort is enormous and challenging.

Other Data Sources

The Governing Board is strongly committed to improving the quality of contextual information while reducing respondent burden and the complexity of data collection and analysis. The self-report questionnaires given to students, teachers, and schools are sometimes burdensome to fill out, labor-intensive to collate and analyze, and subject to concerns about reliability. All questionnaires should be scrutinized to replace as many items as possible with data from centralized records, gathered by test administrators, or, ideally, from computerized data files.

The data available from federal, state, district, and school records should be carefully explored. ~~With implementation of the school report card requirements of the No Child Left Behind law~~, In recent years much more information ~~should be~~has become available ~~soon~~ in standardized computer formats. Barton (2002) has suggested some specific sources of data collected outside of NAEP that should be considered to improve NAEP reporting. These include the U.S. Census, Quality Education Data, Inc. (QED), and the Common Core of Data (CCD) and School and Staffing Survey (SASS), both compiled by the National Center for Education Statistics.

This approach of utilizing more data from outside specific NAEP data collections has been elaborated on extensively in the ~~most recent~~ evaluation of NAEP by the National Academy of Sciences (Pellegrino, J.W., Jones, L.R., & Mitchell, K.J., 1999). The panel proposed "a coordinated system of indicators for assessing educational progress, housed within NCES and including NAEP and other currently discrete, large-scale data collections (p. 34)." Figure 1 is reprinted from the NAS report to show the extent of these data collections on students, teachers, and schools, and to indicate what might be obtained from these other sources. To use them for NAEP would greatly lessen the burden on the assessment itself. Merged data sets could be made available, some to the general public, and more to researchers in restricted data files.

For many years state-level NAEP reports have included appropriate collateral data that provide a context for interpreting NAEP results; see for example the *NAEP 1996 Mathematics: Report Card for the Nation and the States* (Reese et al., 1997). These state contextual variables have included enrollment in elementary and secondary schools, poverty status of children from 5 to 17 years old, number of children receiving disability services, per-pupil expenditures, pupil-teacher ratios, and average teacher salaries. To the extent that these data are readily available and are helpful in setting a context for interpretation of NAEP results the practice ought to be continued. However, more effort should be made to ensure that such data are ~~as~~ up-to-date ~~as possible~~ and easily-accessible as part of NAEP reporting on the Internet.

Figure 1

Overview of Current NCES Data Collections

Data and Design Elements	NAEP	NELS	ELS	ECLS	TIMSS	CCD	PSUS	SASS	NHES
Data Elements									
Student achievement	x	x	x	x	x				
Student background characteristics	x	x	x	x	x	x	x	x	x
Home and community support for learning	x	x		x	x				x
Standards and curricula					x				
Instructional practices and learning resources	x	x		x	x			x	
School organization/governance					x			x	
Teacher education and professional development	x				x	x	x	x	
Financial resources					x	x	x	x	
School climate	x	x		x	x			x	x
Design Elements									
Type of design (CS=cross-sectional; L=longitudinal)	CS,L	L	L	L	CS	L	L	CS,L	CS
Periodicity (TBD=to be determined)	2,4, or 6 yrs	2-6 yrs	TBD	TBD	TBD	Annual	Biennial	2-5 yrs	2-3 yrs
Unit of observation (S=student; T=teacher; A=administrator; P=parent; SC=schools; D=district; ST=states; H=households)	S,T,A	S,T,A	S,A,P	S,T,A,P	S,T,A,P	SC,D,ST	SC	T,A,SC	H
Data collection method (S=survey; R=record analysis; I=interview; V=video; C=case study; O=other)	S	S,R	S,O	S,O	S,R,V,C	S,R	S	S	I
Population of inference (N=national; S=state; G=demographic group)	N,S,G	N,G	N,G	N,G	N	N,S,G	G	N,S,G	N,G

NELS: National Education Longitudinal Study of 1988
ELS: Educational Longitudinal Study of 2002
ECLS: Early Childhood Longitudinal Study
TIMSS: Third International Mathematics and Science Study
CCD: Common Core of Data
PSUS: Private School Universe Survey
SASS: Schools and Staffing Survey
NHES: National Household Education Survey

NOTE: From *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress* (pp.36-37), by J.A. Pellegrino, L.R. Jones, & K.J. Mitchell, 1999, Washington, DC: National Academy Press. Copyright 1999 by the National Academy of Sciences. Reprinted with permission.

NAEP Contextual Information Matrix

The types of descriptive information NAEP collects, reporting units, and data sources are summarized in the NAEP Contextual Information Matrix, which is displayed as Figure 2. The matrix is intended to assist in conceptualizing NAEP contextual information collections. It is organized by priorities—both for types of information and for how data should be obtained. Note that in each case information is to be obtained from reliable official records before it is sought through questionnaires.

The entries in the cells are illustrative, showing the kinds of information that are currently collected by NAEP and the various data sources (records and questionnaires) that are used. As the principles of this framework are implemented, more information will come from records, less from questionnaires. The sources with higher reliability and less respondent burden should be utilized in priority order.

The Ad Hoc Committee on NAEP Background Questions considered a proposal by Paul Barton (2002) to permit states or groups of states to add customized sets of questions to the contextual questionnaires. Although these might track progress on topics of particular interest and increase support for NAEP, the Committee felt strongly that the proposal should not be pursued because any customization of NAEP questionnaires would create serious logistical and quality control problems.

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. The average individual response time to answer contextual questionnaires for each assessment, as calculated in accordance with Office of Management and Budget (OMB) procedures, shall be limited as follows: ~~ten~~ 10 minutes for each student on paper-and-pencil tests, 15-minutes per student on computer-based assessments, 20 minutes for each teacher, and 30 minutes for each school. Consideration should be given to increasing student response time on paper-and-pencil questionnaires if deemed practical and productive.

Figure 2

NAEP Contextual Information Framework

Reporting Unit and Data Sources	Type of Information			
	Student Reporting Categories	**Socio-Economic Status Core| Expanded**	**Other Contextual Information**	**Subject-Specific Information**
STUDENT **School Records** **Questionnaire**	Gender Race/ethnicity SD/LEP Race/ethnicity	Free/RP lunch participation Title I Parent education Reading materials and Internet access in home Own bedroom Parent occupation	New enrollee Type/degree of disability Daily reading Discuss school work ~~TV-watching~~ Absenteeism Language in home After-school learning activities	Course taking in mathematics Time spent on math homework Good in math?
SCHOOL **Dist/State Recds** **School Records** **CCD/Census** **Questionnaire**	School type (public, private, charter, etc.) School ach. data Community type	% Free/RP lunch participation Title I funding	Grade structure Days of instruction Enrollment % LEP % students absent % teachers absent Enrollment mobility Grade retention Teacher retention Graduation rates Post-secondary ed rates	Graduation requirements in math and science. Higher level math courses Graduation testing Extracurricular options in math and English. Availability of computers for writing.
TEACHER **School Records** **Dist/State Recds** **Questionnaire**			Race and Gender Experience Credentials Undergrad/Grad content training Professional Devel	Correct for spelling and grammar? Frequency of science lab work
STATE **CCD/Census** **State Records** **Questionnaire**	Region		Non-NAEP contextual variables	
DISTRICT **CCD/Census** **State Records** **District Records** **Questionnaire**		Community type (urban, rural, etc.)		

NOTE: Information type and data sources are arranged in priority order.

36

Chapter Five: Using Contextual Data to Report NAEP Results

This chapter discusses the descriptive information that NAEP should provide, the levels of disaggregation ~~now~~ possible with merged national and state samples, and the importance of minimizing causal interpretations.

Use of Descriptive Information in NAEP

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on the patterns and trends of non-cognitive variables known to have a relationship to academic achievement.

In addition, special supplemental reports ~~may~~ should be prepared that focus on particular ~~aspects of the background data collected~~topics of public interest and importance. Such reports should feature significant contextual information as well as cognitive results. Advisory committees, including a range of knowledgeable persons, may be appointed to provide input on reporting issues. In all cases, NAEP reports published by the National Center for Education Statistics must not state conclusions as to cause and effect relationships and avoid simplistic presentations unsupported by research that may imply best practice.

All contextual questions and data collected by NAEP should be made available on the Internet at the time of the initial release of the principal academic results or soon afterwards so the public may be able to consider them in discussing results. Complete data files should be available to researchers for further analysis.

~~Implementing~~ No Child Left Behind

The intent of the No Child Left Behind Act of 2001 (P.L.107-110) ~~is~~ has been to hold public schools accountable for closing achievement gaps between different groups of students. NAEP ~~is~~ has~~asked to~~ contributed to this end by providing an accurate measure of current levels of student achievement and ~~to~~ monitoring change over time.

Descriptive information about all students, but particularly on low-performing groups, ~~would~~ can contribute powerfully to the dialogue on the challenges before American education. For example, the NAEP achievement levels focus on the segments of the performance distribution that are at or above *Basic, Proficient*, and *Advanced.* Information ~~should~~ can also be provided about those *Below Basic*, who clearly have been "left behind:" e.g. the proportion having qualified teachers, receiving free or reduced-price lunch, or moving to different schools frequently, as measured by attending the same school for less than two years.

Such profiles of low-performing or high-performing students ~~would~~ should not attempt to ascribe causation, but they ~~would~~ can provide important information on the distribution of practices and resources that are of concern to the public and policy-makers. Periodic collections of such contextual data ~~could~~ can be used to track change in the distribution of these factors over time. Do the trends seem favorable or adverse to educational progress?

Disaggregation of NAEP Data

~~For more than three decades~~Since it was established NAEP has provided data disaggregated by race/ethnicity, gender, school type (e.g., public/private), and community type (e.g., urban/rural). ~~The No Child Left Behind~~Current law calls for disaggregation by major subgroups (when feasible) of race, ethnicity, and gender, and also by socio-economic status, disability, and limited English proficiency.

Because of the large size of the ~~recently~~ combined national and state NAEP samples, NAEP reports ~~should be~~are able to provide information disaggregated at a much greater level of detail than was possible in the program's first decades. ~~Pooling the data from all states, which now are required to provide NAEP samples in fourth and eighth grade reading and mathematics will produce a much enlarged national sample that will sharply reduce the number of empty cells in any cross-tabulations~~. Such disaggregation ~~might~~ adds to the richness of NAEP reporting even ~~with~~ for only a limited set of non-cognitive questions. Disaggregation is also very important for reporting on the distribution of student characteristics within the different achievement levels, as described above.

Minimizing Causal Interpretations

NAEP has often reported on the average performance of students by particular non-cognitive variables. One example, presented in many NAEP reports until the early 2000s, was the average scale score of students ~~who~~ that watch different amounts of television each day, cf. *The Nation's Report Card: Reading, 2000* (Donahue et al., 2001). Another example ~~is~~ has been the average scale scores for 12th graders who report

different amounts of time working at a part-time job, cf. *The Nation's Report Card: Mathematics, 2000* (Braswell et al., 2001).

While there may be a correlation between TV-watching and reading performance, or between hours working outside school and math results, NAEP is not designed to prove cause-and-effect relationships. As a cross-sectional survey, nearly all of its data is on current activities and practices—not on the complex chain of experience in school and outside, of prior learning and achievement that all contribute heavily to current academic performance. While the correlations may be of interest, they cannot be conclusive. But they may be cited to stimulate discussion or encourage further research. ~~Yet, NAEP has encouraged simple causal inferences by reporting average scores for varying amounts of time spent on current activities.~~

There is one important exception to the absence of data on learning-related activity over time. This is the information NAEP collects on the transcripts of high school seniors and its questionnaires on courses that students have taken and schools provide. These do show prior instruction before current exams. The trends in course taking have been of great public interest and it is reasonable to relate them to student achievement.

NAEP reports should present information on the patterns and trends of non-cognitive variables known from other sound research to have a relationship to academic achievement. These presentations should be straightforward and impartial, and care must be taken to avoid stating conclusions as to cause and effect relationships. ~~***Further analysis of any relationships should be left to researchers.***~~

NAEP Data Explorer and Other Online Means of Data Dissemination

Formatted: Indent: First line: 0"

The NAEP Data Explorer should be further improved to make data more accessible to general, non-specialist users. Tables and very simple to construct charts should be prepared to present data on important topics of wide public interest. Additional means of disseminating information through new technology should be explored. These may include simple apps that would allow parents, teachers, and others to access pertinent contextual data as well as NAEP achievement results.

Formatted: Body Text Indent, Left

Chapter Six: Using NAEP in Educational Research

As a cross-sectional survey without longitudinal data, the National Assessment of Educational Progress is able to document school conditions and practices. It can report on achievement results. But it cannot properly be used to establish direct cause-and-effect relationships. Still, over the past ~~three~~ four decades, NAEP has been part of ~~two~~ three important research endeavors—exploring changes in the black-white test score gap since 1970; ~~and~~ seeking to establish the impact of state-level reforms during the 1990s; and evaluating the stringency of state standards enacted under No Child Left Behind.~~-~~

By doing its main task of monitoring achievement well, NAEP has provided sound data for researchers to use. NAEP results have been critical in identifying hypotheses for other research to pursue. Its large data sets, including contextual variables, have been combined with other information to tease out meaning and policy implications, though NAEP's own reports have properly steered clear of these activities.

The Governing Board believes the National Assessment can be of value to educational research and the interest of researchers in the assessment should be encouraged. ~~The NCES program of secondary analysis grants for researchers to use NAEP data should continue~~. Educational researchers should be involved, under the auspices of NCES and its contractors, in developing NAEP contextual questionnaires and other data collection efforts to carry out the provisions of this framework.

This chapter considers the limitations and strengths of NAEP for educational research and discusses research that has made use of NAEP data. The chapter draws on papers by David Grissmer, senior research scientist at RAND, who has used NAEP extensively in analyzing educational factors and trends.

NAEP's Limitations and Strengths for Research

The primary purpose of NAEP is to ***accurately and fairly monitor achievement*** over time ***and accurately and fairly compare achievement across states and important sub-groups of students.*** Beyond providing such data, any research with NAEP, particularly into the causes of academic achievement, is severely limited by its design.

As a representative sample survey, in which no individual student takes more than a small part of the full exam, NAEP has shortcomings in most of the elements commonly used to evaluate academic achievement (Podgursky, 2002):

- It provides no prior data on student achievement, and can't be made longitudinal to do so.

- It can only collect contemporaneous information on school practices and resources, and has no way of ascertaining how students were taught or what school experiences they may have had in previous years.

- There is considerable measurement error in survey responses obtained from teachers and schools because they may well give the expected "right" answers rather than report accurately what they do.

- The current classroom practices that teachers report may be a response to student achievement levels, not the cause of such achievement, and it is difficult to disentangle causation.

- It is difficult for NAEP to get good information on socio-economic status or family background factors, but these are powerfully correlated with academic achievement, and must be controlled for in any analysis of school effects.

On the other hand, NAEP does have unique strengths and comparative advantages (Grissmer, 2003), and thus has the potential to address some important research and public policy questions with its cognitive data and contextual information:

- NAEP is the only data set on student achievement that has collected data from nationally representative samples of students continuously from 1969-70 to the present.

- It is the only data set that has collected academic achievement data simultaneously, repeatedly, and consistently from three separate age groups.

- It is the only data set that collects from statistically reliable samples at the state level, and within states for different types of communities (central city, suburban and rural) and for racial/ethnic groups within most states.

- NAEP has far larger sample sizes than any other nationally representative survey of student achievement, such as the National Education Longitudinal Study (NELS) and the Early Childhood Longitudinal Study (ECLS). These surveys are only approximately 10 to 20 percent as large as NAEP in any single application, and 1 to 5 percent as large as NAEP for any repeated data collection.

- NAEP is the only survey that tests a wide range of academic subjects.

- NAEP achievement measures at fourth and eighth grade fill an important void in measuring the well-being of children during this developmental period.

- NAEP generally incorporates a higher quality and unique design of test instruments, administrative procedures, and scoring methodology, compared to other data sets.

Previous Use of NAEP in Research

As a result of its strengths, NAEP has been used in important educational research by authors such as David Grissmer, Alan Krueger, David Armor, and Christopher Jencks. These studies point to an important comparative advantage of NAEP, namely, that it is the only representative sample data in existence on student achievement in the United States from 1969 to 2002. Thus, research into important historical questions about the effects of changing families, communities, and schools on achievement almost require NAEP data. Without NAEP, it is unlikely that the significant narrowing of the black-white score gap would be known and its possible causes the subject of research.

Similarly, NAEP data have been used to help analyze the effects of differences in resources, systemic reform initiatives, differential opportunity for learning, and other educational policies on state-level academic achievement. Such research has concluded that the rates of improvement in achievement varied markedly across states in the 1990s, and that changing resources or demographics cannot account for the gains in the states with most rapid improvement. This research points to another strong comparative advantage of NAEP. State NAEP is the only survey that includes representative samples of students in many different states, and thus plays a central role in monitoring and explaining the differences in academic achievement and achievement trends across the states. NAEP can identify where positive trends are occurring so researchers can puzzle out causation.

A review of research studies using NAEP (Grissmer, 2003) suggests that only a small proportion of the non-cognitive items collected by the assessment have been utilized in productive research. Also, such research has often supplemented NAEP with data from other sources, such as the U.S. Census and the Common Core of Data (CCD) and Schools and Staffing Survey (SASS), both conducted by the National Center for Education Statistics. However, the National Assessment played such a crucial role in these studies that they could not have been conducted without NAEP data, including some of its non-cognitive variables, principally those on socio-economic status, family structure, and school resources.

On the other hand, NAEP data have also been misused for simplistic and weak research. Many contextual data items on school practice and student behavior have been used in a simplistic way to imply a direct, causal relationship to achievement while ignoring the complex mix of other, more fundamental factors that may well have a stronger impact. NAEP has encouraged such associations by presenting one-way tabulations in its reports, e.g. average scale score by hours of television watched, type of reading instruction, or books read per week, and these have been disseminated widely to support particular beliefs or public policy positions. Simple, single variable linkages can often be misleading because of the strong correlations between many contextual variables, particularly with socio-economic status, prior academic achievement, or family background. They should only be included in NAEP reports when there is strong justification based on previous research.

Also, most of the hundreds of contextual questions in NAEP have never been used for either public reporting or research. Many come from the early 1980s, and would be difficult to justify in a sound research design today.

Secondary Analysis Grants and District Samples

For ~~many years~~almost two decades NCES ~~has been making~~made awards to education researchers for secondary analyses of NAEP data. These explored a range of topics, often in combination with other data sets. Many of the studies ~~have~~ focused on state-to-state differences in student achievement and the impact of state-level policies, relying on NAEP academic data, a few contextual questions for SES controls, and much additional information from other sources. The program ~~has been~~was valuable as a means of encouraging the use of NAEP for research, and, in a few cases, notably the Grissmer studies, ~~has~~ had considerable impact. As in any grant program, all findings are the responsibility of the individual researchers, not of the agency making the grant.

~~The program should continue, and now that~~When NCES ~~has become~~became part of the Institute for Education Sciences in 2003, the ~~leadership of the new agency should ensure that the~~separate NAEP analysis grants were absorbed in a more general research program. ~~are aligned with the research priorities of the Institute.~~We believe this program should increase awards that make use of NAEP data. Efforts should be made through training and other small-scale grants to expand capabilities for using NAEP in productive education research.

In addition, data from the school district NAEP samples in the Trial Urban District Assessment, which started in 2002, ~~will~~ provide important ~~new~~ opportunities for research. NAEP results for school districts can readily be combined with Census data, which include pertinent information on family background and socio-economic status. The school district samples can also be tied to important education policy variables, such as per pupil spending, for which information is available at this level but not for schools.

The primary purpose of NAEP is to provide fair and accurate information on student achievement. Its primary audience is the American public. The Governing Board believes that in serving its purpose and audience well, NAEP can contribute to educational research. It welcomes the interest and efforts of researchers.

Chapter Seven: Review and Improvement of Non-cognitive Questions

This chapter discusses several mechanisms for the review and improvement of NAEP's non-cognitive questions and for implementation of the NAEP Contextual Information Framework.

Independent Validity Studies

Since the early 1990s NAEP has had the benefit of independent outside advice on topics of urgency or interest. These studies have been very helpful to the Governing Board and NCES as they made decisions about the future of the NAEP program. For example, several years ago some research was conducted to examine the possibility of combining the NAEP national and state samples to achieve more efficiency and cost-savings. Starting in 2003 NAEP moved in that direction. The decisions surrounding the change, however, were only as good as the research that bolsters it. The work of the current NAEP Validity Panel, in conjunction with the current NAEP operations contractors, has contributed significantly to making the change possible.

The value of this kind of applied research cannot be overestimated. Neither can the value of the independent nature of this work. The NAEP program is very large and complex and demands a commitment of many resources from the NAEP contractors. NAEP contractors should not be burdened with conducting simultaneous research studies while carrying out the requirements of the operations contracts. There is a precedent for this approach in the current separation of responsibilities for operations and research in separate NAEP contracts. There are two reasons why independent validity studies on topics associated with the non-cognitive framework are recommended. First, there are some non-cognitive variables that will need validation, particularly if those variables are new or are new composite indexes of existing variables. Second, following the approach already established for the NAEP cognitive components, recommendations from such research studies ~~must~~ should be truly independent and free from any conflict of interest.

Review of the Contextual Information Framework

~~This~~ The contextual information framework should be reviewed on a periodic basis. ~~The~~ NAEP cognitive frameworks are reviewed every ten years. This policy was adopted at the time of the NAEP redesign in 1996. Reviewing a NAEP framework can result in major revision, minor revision, or even no revision and re-adoption. The framework may be updated as needed. A thorough review of the Contextual Information Framework should be undertaken ~~Since the background framework is a new undertaking,~~

~~a required review after five years is appropriate with additional reviews~~ every ten years~~.s thereafter.~~

45

References

Allen, N. L., Carlson, J.E., & Zelenak, C.A. (1999). *The NAEP 1996 Technical Report.* Washington, DC: U.S. Government Printing Office.

Armor, D.J. (December 18, 2002). *Comments on NAEP non-cognitive questions.* Available from the National Assessment Governing Board, 800 N. Capitol Street, NW, Washington, DC 20002.

Barton, P.E. (2002, September). *Perspectives on non-cognitive questions in the National Assessment of Educational Progress.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Braswell, J.S., Lutkus, A.D., Grigg, W.S., Santapau, S.L., Tay-Lim, B., & Johnson, M. (2001). *The Nation's report card: Mathematics 2000.* Washington, DC: U.S. Government Printing Office.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of Educational Opportunity.* Report of the Equality of Educational Opportunity Commission (EEOC). Washington, DC: U.S. GPO for the National Center for Education Statistics.

Cowan, C.D., Hauser, R.M. et al (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation. Washington, DC: National Center for Education Statistics.* Available at http://nces.ed.gov/nationsreportcard/researchcenter

Donahue, P.L., Finnegan, R.J., Lutkus, A.D., & Allen, N.L (2001). *The Nation's report card: Reading 2000.* Washington, DC: U.S. Government Printing Office.

Donahue, P.L., Voelkl, K.E., Campbell, J.R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states.* Washington, DC: U.S. Government Printing Office.

Forsyth, R., Hambleton, R.K., Linn, R., Mislevy, R. & Yen, W. (1996). *Design/ feasibility team report to National Assessment Governing Board.* Washington, DC: National Assessment Governing Board. Available at http://nagb.org.

Greenberg, E., Stancavage, F., Farr, B., & Bohrnstedt, G. (January 2001 draft). *Making comparisons: Using background information from the Nation's Report Card to understand student achievement.* Washington, DC: American Institutes for Research.

Grissmer, D. (2002, September). *Improving NAEP for research and policymaking.*

Formatted: Indent: Hanging: 0.5"

Formatted: Font: Not Italic

Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Grissmer, D. (2003, January). *Utilizing NAEP for educational research.* Paper prepared for National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Mayer, D.P., Mullens, J.E., & Moore, M.T. (2001). *Monitoring School Quality: An indicators report.* Washington, DC: U.S. Government Printing Office.

Mullis, I.V.S. (2002, September). *Non-cognitive questions in TIMSS and PIRLS: An overview.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

National Assessment Governing Board. (1996, August). *Policy statement on redesigning the National Assessment of Educational Progress.* Washington, DC: Author.

National Assessment Governing Board. (2002, May). *Policy statement on collection and reporting of non-cognitive data by the National Assessment of Educational Progress.* Washington, DC: Author.

Pellegrino, J.W., Jones. L.R. & Mitchell, K.J. (Eds.) (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* Washington, DC: National Academy Press.

Podgursky, M. (2002, September). *NAEP non-cognitive questions: What can we learn from NAEP about the effect of schools and teachers on student achievement?* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

P.L. 107-110 No Child Left Behind Act of 2001. Section 602. Available at http://nagb.org.

Reckase, M.D. (2002, September). *Contributions of non-cognitive questions to improving the precision of NAEP results.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states.* Washington, DC: U.S. Government Printing Office.

Rust, K. (September 26, 2002). *Technical memo: Sample design for 2003 NAEP.* Available from NAGB Board Meeting briefing book, November 14-16, 2002. Washington, DC. National Assessment Governing Board.

Smith, M.S., Chudowsky, N., Ginsburg, A., Hauser, R.M., Jennings, J., and Lewis, S. (2012). *NAEP Background Questions: An Underused National Resource. A report to the National Assessment Governing Board by the Expert Panel on Strengthening NAEP Background Questions.* Washington, DC. National Assessment Governing Board. Available at http://nagb.org.

Walberg, H.J. (2002, September). *The National Assessment's most useful non-cognitive items.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Formatted: Indent: Hanging: 0.5"

Formatted: Font: Italic

Appendix A

Adopted May 18, 2002

National Assessment Governing Board

Policy Statement on Collection and Reporting of Background Data by the National Assessment of Educational Progress

INTRODUCTION

As the Nation's Report Card, the National Assessment of Educational Progress (NAEP) is an on-going, Congressionally-authorized program to collect data through surveys on the academic knowledge and skills of American students. Its primary goal is to report fair and accurate information on student achievement in reading, mathematics, and other subjects taught in elementary and secondary schools. This information is to be made available in a clear and timely manner to members of the public, policymakers, and educators throughout the country.

Since it began in 1969-70, NAEP has administered, in addition to cognitive questions, background questionnaires that provide information for reporting categories and collect non-cognitive data on students, their family background, teachers, and schools. These have enriched reporting of the National Assessment and increased the precision of NAEP results. The background data have also been used in secondary analyses. However, because NAEP tests a cross-section of students at a particular time with no follow-up of the students tested, the assessment can only show correlations or associations rather than causal relationships between background factors and achievement.

By statute (P.L. 107-110), the National Assessment shall include, "whenever feasible, information collected, cross-tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency." The statute provides that NAEP may "not evaluate or assess personal or family beliefs and attitudes" and may "only collect information that is directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information." These provisions are intended to prevent intrusive, inappropriate, or unnecessary questions being asked about students and their families.

The law requires that the Governing Board take steps to ensure that all NAEP questions are "free from racial, cultural, gender, or regional bias, and are secular, neutral, and non-ideological." However, a House-Senate Conference report, accompanying the legislation, says the law does not preclude the use of "non-intrusive, non-cognitive questions," with a direct relationship to academic achievement.

The National Assessment is conducted by the Commissioner of Education Statistics under the policy guidance of the National Assessment Governing Board. The Board's specific areas of responsibility include: (1) assessment objectives and test specifications; (2) the methodology of the assessment; (3) guidelines for reporting and disseminating results; and (4) "appropriate actions needed to improve the form, content, use, and reporting" of the National Assessment. Under the statute, the Board has "final authority" on the appropriateness of all NAEP items—both cognitive and non-cognitive.

To carry out these responsibilities, the National Assessment Governing Board hereby adopts guiding principles, policies, and procedures for the collection and reporting of background data by the National Assessment of Educational Progress.

GUIDING PRINCIPLES

1. Background data on students, teachers, and schools is needed to fulfill the statutory requirement that NAEP include information, whenever feasible, disaggregated by race or ethnicity, socioeconomic status, gender, disability, and limited English proficiency. In addition, background data is collected to enrich the reporting of NAEP results by examining factors related to academic achievement. However, the collection of such data should be limited, and the burden on respondents kept to a minimum. It must always be considered in light of NAEP's primary purpose: providing sound, timely information on the academic achievement of American students.

2. All background questions must be directly related to academic achievement or to the fair and accurate presentation of achievement results.

3. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining the background questions to be asked and the samples to which they shall be administered.

4. In accordance with law, questions shall be non-intrusive and free from bias, and must be secular, neutral, and non-ideological.

5. No personally identifiable information shall be included in NAEP reports or data releases.
6. Decisions on the retention or addition of background items shall draw on technical studies on the reliability and validity of current and proposed

questions and on the contribution such items make to the precision of NAEP results.

7. Consideration should be given to obtaining background information from non-NAEP sources and to avoiding duplication with other federal surveys.

8. Questionnaires should be revised to keep background questions timely and related to academic achievement. Those questions showing little change over time and/or a stable relationship to achievement should be deleted or asked less frequently and to limited samples, unless required to assure the precision of NAEP results.

9. Questions should not address personal feelings and attitudes.

10. Since security considerations do not apply, background questionnaires shall be readily available to the public.

11. Interpretation of results shall be limited in official NAEP reports and must be strongly supported by NAEP data. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships.

12. Background questions for NAEP assessments shall be prepared in accordance with frameworks and specifications adopted by the Governing Board.

13. The Governing Board shall review and approve all background items before they are administered in NAEP surveys or pilot and field tests.

POLICIES AND PROCEDURES

1. Framework and Specifications

The Governing Board will adopt a general framework for background questionnaires and specifications for the questions on selected topics and in specific subject areas.

Since this is a new area of responsibility for the Board, the process of developing a framework for background questions and specifications will begin with commissioned papers on relevant issues, such as the reliability and validity of current background questions, their contribution to improving the precision of NAEP results, their value and limitations for educational research, and changes that may be needed in response to the No Child Left Behind legislation. Following consideration of these issues, the Board will define the scope of background questionnaires and adopt a process for preparing a framework and specifications. This work will include the active participation of teachers,

education researchers, state and local school administrators, assessment specialists, parents of children in elementary and secondary schools, and interested members of the public.

2. Background Question Development

In preparing background questions, the National Center for Education Statistics shall follow adopted frameworks and specifications, and consider the review criteria adopted by the Governing Board. NCES may use cognitive laboratories of students, teachers, and school officials to help determine the clarity and burden of proposed questions. Ad hoc advisory committees may also be established, comprised of teachers, parents, technical experts, and others interested in NAEP. Steps shall be taken to determine the reliability of questions used.

3. Governing Board Review and Approval of Background Questions

Background questions for all NAEP pilot tests, field tests, and operational use shall be reviewed and approved by the Governing Board. The category of respondents, e.g. students, schools, and grade level, shall clearly be designated, as will the NAEP samples, e.g. national, state, or district, in which the questions will be asked.

For each questionnaire there shall be an explanation of its intended use in NAEP reporting and analysis and of the hypothesized relationships between the background items and student achievement that demonstrates the need to know such information. Technical data shall be presented on the reliability and validity of questions and, if applicable, on their contribution to improving the precision of NAEP results. The Board will use the explanations and data presented along with the review criteria in this policy statement in determining the appropriateness of background questions.

The Reporting and Dissemination Committee will have primary responsibility for the review and approval of background questions. The Assessment Development Committee will participate in the approval of questions relating to specific subject-matter assessments. Ad hoc committees of Board members may be established by the Board Chairman for background question review. Questions may also be reviewed by external advisors, including teachers, parents, and technical experts. Recommendations on background questionnaires shall be subject to final approval by the full Governing Board.

4. Criteria for Governing Board Review

The following criteria for review and approval of background questions are based on the most recent revision of the authorizing statute of the National Assessment of Educational Progress (P.L. 107-110) and the Guiding Principles of this policy statement:

A. Background information is needed to fulfill the statutory requirement that NAEP report and analyze achievement data, whenever feasible, disaggregated by race or ethnicity, gender, socio-economic status, disability, and limited English proficiency. Non-cognitive data may enrich the reporting and analysis of academic results, but the collection of such data should be limited and the burden on respondents kept to a minimum.

B. All background questions must be related to the primary purpose of NAEP: the fair and accurate presentation of academic achievement results.

C. Any questions on conditions beyond the school must be non-intrusive and focused on academic achievement and related factors.

D. Questions shall be free from racial, cultural, gender, or regional bias.

E. All questions must be secular, neutral, and non-ideological. Definitions of these terms, accompanied by clarifying examples, are presented in Appendix A, as adopted in the Governing Board Policy on NAEP Item Development and Review.

F. NAEP must not evaluate or assess personal feelings or family beliefs and attitudes unless such questions are non-intrusive and have a demonstrated relationship to academic achievement.

G. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining which questions to include in background questionnaires. These factors must also be considered in determining the frequency with which various questions shall be administered and whether they shall be included in both national and state samples.

H. Background questions that do not differentiate between students or have shown little change over time should be deleted or asked less frequently and to limited samples.

5. Public Access to Background Questions

Since security considerations do not apply, all background questionnaires shall be readily available to parents, teachers, state and local officials, and interested members of the public. Such questionnaires shall be available before

field tests and operational assessments or at any other time members of the public wish to obtain them. Background questions in operational use shall be posted on the Internet prior to each assessment, accompanied by explanations and rationales.

6. Reporting of Background Information

The presentation of background data in official NAEP reports shall be straightforward and impartial. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships. Any composite indices including demographic and socioeconomic factors shall be presented to the Board for approval before use as reporting categories in NAEP data releases and reports.

Background data should be available for extensive secondary analyses by scholars and researchers, who are responsible for conclusions reached. Responses to background questions shall be presented and tabulated on the Internet, although, if necessary, posting may be delayed for a brief period after release of the principal NAEP results.

Definitions of Secular, Neutral, and Non-ideological Item Review Criteria

From Governing Board Policy on NAEP Item Development and Review— 5/18/02

Items shall be secular, neutral, and non-ideological. Neither NAEP nor its questions shall advocate a particular religious belief or political stance. Where appropriate, NAEP questions may deal with religious and political issues in a fair and objective way. The following definitions shall apply to the review of all NAEP test questions, reading passages, and supplementary materials used in the assessment:

Secular — NAEP questions will not contain language that advocates or opposes any particular religious views or beliefs, nor will items compare one religion unfavorably to another. However, items may contain references to religions, religious symbolism, or members of religious groups where appropriate.

Examples: The following phrases would be acceptable: "shaped like a Christmas tree," "religious tolerance is one of the key aspects of a free society," "Dr. Martin Luther King, Jr. was a Baptist minister," or "Hinduism is the predominant religion in India."

Neutral and Non-ideological — Items will not advocate for a particular political party or partisan issue, for any specific legislative or electoral result, or for a

single perspective on a controversial issue. An item may ask students to explain both sides of a debate, or it may ask them to analyze an issue, or to explain the arguments of proponents or opponents, without requiring students to endorse personally the position they are describing. Item writers should have the flexibility to develop questions that measure important knowledge and skills without requiring both pro and con responses to every item.

Examples: Students may be asked to compare and contrast positions on states rights, based on excerpts from speeches by X and Y; to analyze the themes of Franklin D. Roosevelt's first and second inaugural addresses; to identify the purpose of the Monroe Doctrine; or to select a position on the issue of suburban growth and cite evidence to support this position. Or, students may be asked to provide arguments either for or against Woodrow Wilson's decision to enter World War I. A NAEP question could ask students to summarize the dissenting opinion in a landmark Supreme Court case.

The criteria of neutral and non-ideological also pertain to decisions about the pool of test questions in a subject area, taken as a whole. The Board shall review the entire item pool for a subject area to ensure that it is balanced in terms of the perspectives and issues presented.

Appendix B

Adopted August 4, 2012

Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting

INTRODUCTION

By statute, the purpose of the National Assessment of Educational Progress is to provide a "fair and accurate" measure of student achievement and achievement trends. Academic or cognitive questions are its primary focus; the American public is its primary audience. However, in addition to reporting on what American students know and can do, NAEP has collected data for more than 40 years that provide a context for reporting and interpreting achievement results. According to the statute, such factors, both in and out of school, must be "directly related to the appraisal of academic achievement."

In each assessment NAEP administers background questionnaires for students, their teachers, and schools. The questionnaires deal with educational experiences and other factors, such as teacher training or out-of-school learning activities, that are related to academic achievement. Data on several hundred background or noncognitive variables are available on the Internet through the NAEP Data Explorer. However, for more than a decade, little use has been made of this information in NAEP reports. The data have received minimal attention and had little impact despite the considerable efforts expended in developing and approving questionnaires and collecting and tabulating responses.

In October 2011 the National Assessment Governing Board convened an expert panel to recommend how to make better use of existing NAEP background questions and to propose an analytic agenda for additional topics and questions that would be useful in developing education policy and of value to the public. The panel report, entitled, *NAEP Background Questions: An Underused National Resource*, was presented to the Board in March 2012 by Marshall Smith, former U.S. Under Secretary of Education, who chaired the six-member panel.

Many of the panel recommendations build on the *Background Information Framework for the National Assessment of Educational Progress*, adopted by the Governing Board after it received final authority from Congress over non-cognitive items on the assessment. The framework was adopted in 2003, but has not been fully implemented.

The following policies are based on recommendations by the expert panel. The Board has also taken into consideration a wide range of public comment and the analysis provided by the National Center for Education Statistics.

It is important to understand that the National Assessment is not designed to show cause-and-effect relationships. Its data should not be used to "prove" what schools should do. But, as the *Background Information Framework* declares, NAEP's "descriptions of the educational circumstances of students…, considered in light of research from other sources, may provide important information for public discussion and policy action." The Board believes the National Assessment should improve upon its efforts to collect contextual information and present it clearly to the public, which will add to NAEP's value to the nation.

POLICY PRINCIPLES

1. NAEP reporting should be enriched by greater use of contextual data derived from background or non-cognitive questions asked of students, teachers, and schools. Such data will be used both in regular Report Cards and in special focused reports.

2. Reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Care should be taken not to suggest causation.

3. Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in background questionnaires and their connection to student achievement. Such frameworks should be updated for each assessment cycle and provide the basis for new topics and questions.

4. An ad hoc committee of the Board will be established for one year to monitor implementation of this resolution, review the *NAEP Background Information Framework*, and recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP.

IMPLEMENTATION GUIDELINES

For Questions and Questionnaires

1. Clusters of questions will be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning, which could be used regularly or rotated across assessment cycles.

2. Modules will be prepared for special one-time studies to provide descriptive information on issues of current policy interest.

3. A thorough review will be conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions will be dropped.

4. NAEP will include background questions from international assessments, such as PISA and TIMSS, to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

5. Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest.

6. The development and use of improved measures of socio-economic status (SES) will be accelerated, including further exploration of an SES index for NAEP reporting.

For Data Collection

7. The maximum time for students to answer the background questionnaire will be increased from 10 to 15 minutes on new computer-based assessments. Consideration should be given to a similar increase in paper-and-pencil assessments.

8. Whenever feasible, assessment samples should be divided (spiral sampling) and background questions rotated in different years in order to cover more topics without increasing respondent burden. These practices will be initiated in the assessments of reading and mathematics, which are conducted frequently, and considered for other subject areas if the frequency of testing permits.

For Reporting

9. Special focused reports with data through the 2013 assessment will be issued on the following topics: private schools, charter schools, gender gaps, and black male students. Reports shall include significant contextual information as well as

cognitive results. Advisory committees, composed of a range of knowledgeable persons, may be appointed to provide input on reporting issues.

10. Exploratory analyses will be carried out to determine if existing background questions may form the basis for additional focused reports. Such reports may be issued by the Governing Board as well as by the National Center for Education Statistics.

11. The NAEP Data Explorer should be further improved to make data more accessible to general, non-specialist users. Tables and very simple-to-construct charts will be prepared to present data on important topics of wide public interest. Additional means of disseminating information, using new technology such as simple apps that would allow parents, teachers, and others to access background and achievement data, will be explored.

APPENDIX A
Acknowledgements

The Ad Hoc Committee on NAEP Background Questions, convened in 2002, was chaired by Governing Board member John H. Stevens. He also served as chairman of the Board's standing Committee on Reporting and Dissemination, which has responsibility for reviewing all core NAEP background questionnaires and making recommendations on them to the full Board.

The Ad Hoc Committee also included members of the Board's two other standing Committees—Assessment Development and Standards, Design, and Methodology—with a wide range of backgrounds and perspectives: Amanda Avallone, Dwight Evans, Thomas Fisher, Sheila Ford, Jo Ann Pottorff, and Sister Lourdes Sheehan. The Board Chairman, Darvin Winick, participated in many of the lively discussions that marked the Committee's deliberations.

Among the many discussants and presenters at the workshop and public forum, we wish to recognize the care and thoughtfulness of Robert Mislevy, of the University of Maryland, and Harold Wenglinsky, of Baruch College of the City University of New York. The comments submitted by Sandra Feldman, president of the American Federation of Teachers, were particularly perceptive and helpful. This project also benefited greatly from the continuing advice and insight of Paul Barton and David Grissmer, both of whom have used NAEP data for many years to understand and explain American education to its public.

The Ad Hoc Committee convened in 2012 was chaired by Terry Holliday and also included seven other Governing Board members: Doris Hicks, Andrew Ho, Brent Houston, Dale Nowlin, Joseph O'Keefe, S.J., Susan Pimentel, and Leticia Van de Putte. Lawrence Feinberg, Assistant Director for Reporting and Analysis, provided staff support to both the 2002 and 2012 committees.

ADOPTED UNANIMOUSLY BY NAGB - 8/4/12 with **follow-up in revised framework**

Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting

INTRODUCTION

By statute, the purpose of the National Assessment of Educational Progress is to provide a "fair and accurate" measure of student achievement and achievement trends. Academic or cognitive questions are its primary focus; the American public is its primary audience. However, in addition to reporting on what American students know and can do, NAEP has collected data for more than 40 years that provide a context for reporting and interpreting achievement results. According to the statute, such factors, both in and out of school, must be "directly related to the appraisal of academic achievement."

In each assessment NAEP administers background questionnaires for students, their teachers, and schools. The questionnaires deal with educational experiences and other factors, such as teacher training or out-of-school learning activities, that are related to academic achievement. Data on several hundred background or noncognitive variables are available on the Internet through the NAEP Data Explorer. However, for more than a decade, little use has been made of this information in NAEP reports. The data have received minimal attention and had little impact despite the considerable efforts expended in developing and approving questionnaires and collecting and tabulating responses.

In October 2011 the National Assessment Governing Board convened an expert panel to recommend how to make better use of existing NAEP background questions and to propose an analytic agenda for additional topics and questions that would be useful in developing education policy and of value to the public. The panel report, entitled, *NAEP Background Questions: An Underused National Resource*, was presented to the Board in March 2012 by Marshall Smith, former U.S. Under Secretary of Education, who chaired the six-member panel.

Many of the panel recommendations build on the *Background Information Framework for the National Assessment of Educational Progress*, adopted by the Governing Board after it received final authority from Congress over non-cognitive items on the assessment. The framework was adopted in 2003, but has not been fully implemented.

The following policies are based on recommendations by the expert panel. The Board has also taken into consideration a wide range of public comment and the analysis provided by the National Center for Education Statistics.

It is important to understand that the National Assessment is not designed to show cause-and-effect relationships. Its data should not be used to "prove" what schools should do. But, as the *Background Information Framework* declares, NAEP's "descriptions of the educational circumstances of students…, considered in light of research from other sources, may provide important information for public discussion and policy action." The Board believes the National Assessment should improve upon its efforts to collect contextual information and present it clearly to the public, which will add to NAEP's value to the nation.

POLICY PRINCIPLES

1. NAEP reporting should be enriched by greater use of contextual data derived from background or non-cognitive questions asked of students, teachers, and schools. Such data will be used both in regular Report Cards and in special focused reports.
[New Foreword, pp. 7 and 9]

2. Reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Care should be taken not to suggest causation. **[Chapter 5, pp. 37-39; also pp. 7 and 9]**

3. Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in background questionnaires and their connection to student achievement. Such frameworks should be updated for each assessment cycle and provide the basis for new topics and questions. **[p. 13]**

4. An ad hoc committee of the Board will be established for one year to monitor implementation of this resolution, review the *NAEP Background Information Framework*, and recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP.

IMPLEMENTATION GUIDELINES

For Questions and Questionnaires

1. Clusters of questions will be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning, which could be used regularly or rotated across assessment cycles.
[pp. 13 and 30]

2. Modules will be prepared for special one-time studies to provide descriptive information on issues of current policy interest. **[p.29]**

3. A thorough review will be conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions will be dropped. **[p. 13]**

4. NAEP will include background questions from international assessments, such as PISA and TIMSS, to obtain direct comparisons of states and TUDA districts to educational practices in other countries. **[pp. 12 and 23]**

5. Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest. **[pp. 11 and 23]**

6. The development and use of improved measures of socio-economic status (SES) will be accelerated, including further exploration of an SES index for NAEP reporting. **[pp. 12 and 27]**

For Data Collection

7. The maximum time for students to answer the background questionnaire will be increased from 10 to 15 minutes on new computer-based assessments. Consideration should be given to a similar increase in paper-and-pencil assessments. **[pp. 13 and 35]**

8. Whenever feasible, assessment samples should be divided (spiral sampling) and background questions rotated in different years in order to cover more topics without increasing respondent burden. These practices will be initiated in the assessments of reading and mathematics, which are conducted frequently, and considered for other subject areas if the frequency of testing permits. **[pp. 13 and 30]**

For Reporting

9. Special focused reports with data through the 2013 assessment will be issued on the following topics: private schools, charter schools, gender gaps, and black male students. Reports shall include significant contextual information as well as cognitive results. Advisory committees, composed of a range of knowledgeable persons, may be appointed to provide input on reporting issues. **[p. 37]**

10. Exploratory analyses will be carried out to determine if existing background questions may form the basis for additional focused reports. Such reports may be issued by the Governing Board as well as by the National Center for Education Statistics.

11. The NAEP Data Explorer should be further improved to make data more accessible to general, non-specialist users. Tables and very simple-to-construct charts will be prepared to present data on important topics of wide public interest. Additional means of disseminating information, using new technology such as simple apps that would allow parents, teachers, and others to access background and achievement data, will be explored. **[p. 39]**

Using NAEP Data for Key Education Indicators

As authorized by the Governing Board Policy Statement on NAEP Background Data adopted in 2012, consultants have been preparing an exploratory analysis on using NAEP data for key education indicators. The purpose of this project is to illustrate the usefulness of NAEP in developing a limited number of indicators to represent crucial components of the education system and their interrelationships. The key idea is that instead of starting with background variables and looking for education issues they might address, there should first be a framework of important education policy issues and objectives that can be used to identify relevant background variables.

The work is being undertaken by Alan Ginsburg, former Director of Policy and Program Evaluation at the U.S. Department of Education, and Marshall (Mike) S. Smith, former U.S. Under Secretary of Education and former Dean of the Stanford University Graduate School of Education. Smith chaired the Board's Expert Panel on Strengthening NAEP Background Questions, which presented its report in February 2012. Ginsburg served as a panel member and executive secretary, and has prepared several other exploratory analyses for the Board.

As explained in the statement of work for the project, an education indicator is an individual or composite statistic that measures progress toward an educational objective and is useful in a policy context. Such objectives are concerned not only with student performance but with the quality, equity, and efficiency of the education system in supporting academic achievement. One possible indicator might be the percentage of 8th grade science students with a teacher who majored or minored in science in college. Others might be the extent of severe absenteeism or the use of technology in science instruction.

At the joint meeting Alan Ginsburg and Mike Smith will discuss the purposes and outcomes of the indicator project and how organizing contextual questionnaires to produce data for education indicators might increase the usefulness and impact of NAEP.

DEVELOPING A NAEP INDICATORS FRAMEWORK: LESSONS FROM MAJOR INTERNATIONAL AND DOMESTIC EDUCATION INDICATOR REPORTS

EXECUTIVE SUMMARY

By Alan Ginsburg and Marshall S. Smith

Introduction

This is the first of two reports exploring the use of background data collected by the National Assessment of Educational Progress (NAEP) to develop *key education indicators* at national, state, and urban district levels. *Key indicators are statistics that regularly measure an important condition of education.* For example, NAEP can tie to its achievement results the reporting of background conditions on: student attitudes toward learning, motivation and excessive absenteeism; measures of teacher quality; and indicators of the nature of reading and math instruction (e.g., instructional time).

The Government Accountability Office identified three broad purposes of indicators:

- Increase transparency and public awareness.
- Foster civic engagement and collaboration.
- Monitor progress, establish accountability for results, and aid decision-making.

In a NAEP context, indicators would also serve to:

- Identify for each subject assessed (e.g. reading) a set of key indicators, which are derived from the background variables and are continuously monitored.

Specifically, this first report is intended to develop a general indicators framework specifying an organizing structure, potential indicators, measurement criteria and reporting design. The report is based on a review of several major international and domestic data collections and reports produced by organizations other than NAEP:

International

- Organization for Economic Cooperation and Development, *Education At a Glance*
- International Education Association's 2011 TIMSS *Mathematics Assessment* covering grade 4 and 8.
- OECD's 2009 *PISA Report*

Domestic

- National Center for Education Statistics, *Condition of Education*
- Education Week's *Quality Counts*
- U.S. Department of Education's *Annual Priority Performance Goals*
- National Academy of Sciences' *Key National Education Indicators*

Potential Indicators by Organizing Structure

Exhibit EX-1

Potential NAEP Education Indicators From Which To Select Key Indicators For K-12			
Locus of Education Activity	**Key Drivers** - 21st Century Skills - Common Core Standards - Instructional Technology		
	Results	**Enablers**	**Context/Constraints**
Student	• Command of core content, using NAEP scores • College readiness levels by age and grade • Career readiness (21st century skills)	• Attended preprimary education • Chronic absenteeism • Student motivation and belief that hard work is more important than luck • Student positive attitudes toward subject • Student uses research-based approaches to learning subject • Student respect for teacher and visa versa • Participation in extra-curricular activities including community service	• Home learning environment • Formal and informal learning outside school – nature of the their neighborhood
Teacher	• Proportion of teacher evaluations that distinguish them from a basic standard • Quality of work that the students have • Teachers spend time supporting other teachers	• Teachers with less than 3-years experience • Teachers with mastery-level and current knowledge of content they are teaching. • Teachers with mastery-level and cotemporary knowledge of child and adolescent development • Teacher-student interactions that demonstrate high levels and qualities of involvement, language, stimulation, and expansion of thinking and cognition, and sensitivity to students' perspectives, individual experiences, and backgrounds • Teacher student interactions that indicate that teachers respect students.	• Teacher working conditions • Average district teacher salary • Time teachers spend teaching • Teacher has high quality professional development and comprehensive induction programs Quality of the principal • Teachers belong to professional learning communities
School/ Classroom	• School subject area assessment outcomes • School performance rating/ranking within their state • Parent satisfaction (on surveys) • Completion rates from each kind of school – elementary to middle, middle to high, high to graduate, graduate to college or job?	• Content of instruction aligned with standards • Effective use of technology to support instruction • School Climate – whether the school is a learning organization – do teachers work together? • Instructional time per subject • Engaged instruction in subject • Emphasis on continuous improvement on outcomes through both formative and summative assessments aligned with standards • Emphasis on continuous improvement of practices of teaching	• School SES Composition • Safe & orderly school climate • Teacher-student ratio • School resource shortages • School lacks key characteristics, coaches for teachers, support systems for students, technology, books
System (district, state or nation)	- System core content outcomes	• Support for implementation of new content standards • Alignment of assessment with content standards • Accountability with emphasis on continuous improvement	• K-12 education spending as a share of gross domestic product • K-12 spending per student • Disparity in resources across districts within states

The indicator structure in Exhibit EX-1 is focused primarily around variables at student, teacher, school/classroom and system levels that support learning outcomes across three aspects of education conditions:

- *Results* indicators include student assessment outcomes (such as from NAEP), but also teacher evaluations that reflect student outcomes, and other outcomes such as secondary school completion and parent satisfaction with the school.

- *Enabler indicators* reflect formal learning at different levels of education. These include student exposure to preschool, teachers' knowledge and skills and their ability to apply them to create a challenging and supportive classroom learning environment; and school instructional time and student engagement in the content areas. Enablers also include system policies and regulations at district, state and national levels regarding teacher certification, standards, assessment, and accountability.

- *Context/constraint indicators* reflect factors not readily manipulable by educators but may be changed by policy and funding shifts or proper interventions in the home learning environment. These factors include: learning at home and outside the school in formal and informal settings; factors influencing teacher quality, such as salaries and working conditions; and factors affecting the school learning environment including school safety, climate and class size.

Indicator Measurement

A sound measure for an indicator should meet criteria of validity, reliability, and consistency overtime.

Validity of Indicators. A valid measure is one that adequately captures the underlying education condition of interest. Combining responses from a number of questions around a topic into a larger comprehensive indicator scale produces richer indicator measures than reporting on a single question, but this approach currently is not used in NAEP background factor analyses. Exhibit EX-2 illustrates a scale developed from TIMSS at grade 4 measuring students' early numeracy activities before beginning primary school.

Exhibit EX-2 Development of Indicator Scales from Multiple Questions

Exhibit 4.9: Early Numeracy Activities Before Beginning Primary School*

TIMSS 2011 Mathematics 4th Grade

Reported by Parents

Students were scored according to their parents' frequency of doing the six activities on the *Early Numeracy Activities* scale. Students **Often** engaged in early numeracy activities had a score on the scale of at least 10.3, which corresponds to their parents "often" doing three of the six activities with them and "sometimes" doing the other three, on average. Students **Never or Almost Never** engaged in such activities had a score no higher than 6.9, which corresponds to parents "never or almost never" doing three of the six activities with them and "sometimes" doing the other three, on average. All other students had parents who **Sometimes** engaged them in early numeracy activities.

Country	Often		Sometimes		Never or Almost Never		Average Scale Score
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	
Hungary	75 (0.9)	528 (2.9)	23 (0.8)	495 (4.9)	1 (0.4)	~ ~	11.1 (0.04)
Czech Republic	75 (0.8)	514 (2.3)	25 (0.8)	508 (3.6)	0 (0.1)	~ ~	11.0 (0.03)
Slovak Republic	73 (1.0)	514 (3.3)	25 (0.8)	499 (5.2)	2 (0.5)	~ ~	11.1 (0.05)
Northern Ireland s	70 (1.2)	583 (3.5)	29 (1.2)	566 (4.9)	1 (0.2)	~ ~	11.2 (0.05)
Russian Federation	69 (1.1)	547 (3.7)	29 (1.0)	533 (4.7)	2 (0.3)	~ ~	10.9 (0.04)
Poland	68 (0.9)	488 (2.3)	31 (0.9)	471 (3.0)	1 (0.2)	~ ~	10.8 (0.03)
Ireland	66 (0.9)	539 (2.9)	33 (0.8)	517 (3.7)	2 (0.3)	~ ~	10.9 (0.04)
Australia s	61 (1.2)	540 (3.7)	36 (1.2)	520 (4.1)	3 (0.4)	488 (13.4)	10.7 (0.05)
Austria	61 (0.9)	515 (2.6)	38 (0.9)	502 (3.5)	2 (0.2)	~ ~	10.4 (0.03)
Croatia	60 (0.8)	496 (2.3)	39 (0.8)	482 (2.6)	1 (0.2)	~ ~	10.5 (0.03)
Germany r	59 (1.0)	538 (2.3)	40 (1.0)	528 (2.8)	2 (0.2)	~ ~	10.4 (0.03)

Source: IEA, *TIMSS,* 2011

Reliability of Indicators. A reliable indicator produces consistent results when repeatedly measuring the same underlying condition. Qualitative responses may be unreliable when sensitive to the position of the respondent. For example, Exhibit EX-3, taken from the NAEP background paper on science (by Alan Friedman and Alan Ginsburg), shows that teachers were more likely to indicate that resources within a school are "not at all available" than were principals in the same school. This is not surprising as it is principals who are responsible for school resource availability.

Exhibit EX-3 Differences between teacher and school reported responses about science resource availability raise issues of response reliability

	Not at all	Small Extent	Moderate extent	Large extent
Science Kits are provided (teacher reported)	26	30	29	16
Science Kits are provided (school reported)	7	24	32	37
Science magazines and books are provided (teacher reported)	22	35	33	11
Science magazines and books are provided (school reported)	2	19	35	44

NAEP Data Explorer

Consistency over time. A consistent measure requires using the same measure for an indicator over time. When measures are changed from time period to time period it is unclear whether a change comes about because of a real change in the underlying condition or changes in the measure. The report by the Expert Panel on Strengthening NAEP Background Questions (2012) addressed this issue in its recommendation 1d:

> "Use consistency over time as a criterion to consider for question selection and wording. NAEP's inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.

For example, the Expert Panel found that only one-third of the 2011 questions asking about course offerings provided at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

Sources of Indicator Data. The reports that were studied use two ways to obtain indicator data, which differentiate them from NAEP.

First, TIMSS and PISA both conduct a household survey to obtain information directly from parents or guardians about socio-economic status and the home learning environment. TIMSS innovatively combined with PIRLS to develop a joint sample household survey for grade 4 students. The household survey included questions about:

- Early numeracy activities in the home before beginning primary school (See Exhibit EX-2)
- Early literacy activities in the home before beginning primary school
- Amount of exposure to preschool
- Family perception about child's literacy and numeracy skills before entering primary school
- Family interaction with the child about school work
- Family perceptions about school
- Family literacy environment
- Family SES

A second source of data that is different from NAEP is the pooling of information across different surveys. The Condition of Education and Education at a Glance are drawn almost entirely from data series generated by other surveys. Quality Counts is a state-level amalgam of direct analyses of state policies by Education Week combined with data from other surveys, which prominently features NAEP assessment results.

A form of pooling could be the aligning of NAEP survey questions with international assessment items as illustrated in Exhibit EX-4. The exhibit suggests that at least for U.S. middle schools, only about 12 percent of U.S. principals are having at least some difficulty filling vacancies for mathematics teachers. This compares with other Western English-speaking countries of 41 percent of the principals having difficulty hiring math teachers in Australia, 37 percent in England, and 44 percent in New Zealand. Adding

the same question about vacancies to the NAEP principal survey for mathematics would yield U.S. state-by-state comparisons.

Exhibit EX-4

Schools Having Difficulties Filling Vacancies With Mathematics Teachers, Grade 8

Country	No Vacancies	Vacancies Are Easy To Fill	Vacancies Are Somewhat Difficult To Fill	Vacancies Are Very Difficult to Fill	Total of Vacancies Somewhat or Very Difficut To Fill
	Percent of students	Percent of students	Percent of students	Percent of students	Percent of students
Australia	25	34	31	10	41
Chinese Taipei	46	44	10	1	11
England	28	35	27	10	37
Finland	42	46	10	1	11
Hong Kong SAR	48	44	8	0	8
Japan	82	6	8	3	11
Korea, Rep. of	67	16	15	2	17
New Zealand	30	27	38	6	44
Norway	38	40	20	2	22
Russian Federatio	81	11	6	2	8
Singapore	59	38	2	0	2
United States	63	25	9	3	12

Source : 2011 TIMSS, Mathematics

Next Steps: Using the International and Domestic Indicator Framework as a Guide, Develop a NAEP Education Indicators Framework and Provide Examples with Current Data

A second report will be prepared for the Governing Board with a recommended set of *Key Indicators* and recommended improvements in NAEP data to strengthen indicator measurement or fill indicator gaps. This report will:

- Specify a NAEP Education Indicators Framework for the background variables applicable across cognitive assessments.
- Propose indicators that are research-based and estimable by:
  - offering examples using current NAEP data.
  - suggesting changes to the current NAEP questionnaires.
  - introducing a fundamentally new NAEP questionnaire or drawing data from education surveys other than NAEP.
- Explore opportunities for combining NAEP with international or other NCES indicator-supporting data.
- Explore how NAEP reports could best display a pyramid information approach along the lines of an indicator dashboard.

MARSHALL S. SMITH

Marshall (Mike) S. Smith is retired and a Senior Fellow in Education Policy at the Carnegie Foundation for the Advancement of Teaching. He is a board member of a number of non-profit organizations in the San Francisco Bay Area. During the first two years of the Obama administration he served as Senior Counselor to Secretary of Education Arne Duncan and as Director of International Affairs. From 2001-2009 he directed the Education Program at the William and Flora Hewlett Foundation where he focused on developing the Open Education Resources movement, improving instruction, and reforming California's educational system.

Prior to that, in the Clinton Administration, he was the Undersecretary of Education for seven years responsible for all policy and budget matters. For the last four of those years he also served as the acting deputy secretary, the Education Department's second-ranked official under Secretary Richard Riley. During the Carter administration, he served as chief of staff to the first secretary for education, Shirley Hufstedler, and assistant commissioner for policy studies in the Office of Education. In the Ford administration he was the director of policy and budget for the National Institute of Education, the education research arm of the U.S. Government. While not in government, he was at different times an associate professor at Harvard and a professor at the University of Wisconsin (at Madison) and at Stanford University. At Stanford, he was also the dean of the School of Education.

Smith has authored a large number of publications on topics varying from computer content analysis, evaluation and research methodology, social and educational inequality, early childhood education, open educational resources, federal policy, standards-based reforms and the use of technology in education in the developed and developing worlds. He is a member of the National Academy of Education and a fellow of the American Academy of Arts and Sciences. He holds bachelor's, master's, and doctorate degrees from Harvard.

ALAN GINSBURG

Alan Ginsburg was Director of Policy and Program Studies (retired) within the Office of Planning, Evaluation, and Policy Development at the U.S. Department of Education. He coordinated the Education Department's Government and Performance Results Act indicators and annual reports to Congress. Ginsburg's international work includes: Lead Shepherd (chair) of the Human Resources Development Working Group; and chair of the APEC Education Network (EDNET). His international mathematics work about Singapore and other Asian countries is extensively cited by the Common Core State Standards Initiative.

Ginsburg received his Ph.D. in economics from the University of Michigan. He received the Distinguished Presidential Rank Service Award, the federal government's highest award given to its civil service employees. He also received the American Evaluation Association's Gunnar Myrdal award for his contributions to the field of evaluation. He has been advisor to Education Week on their annual reports.

Upcoming NAEP Reports as of December 2013

Report	Expected Release Date

Initial NAEP Releases

2013 Mathematics TUDA Grades 4, 8	December 2013
2013 Reading TUDA Grades 4, 8	December 2013
2013 Reading and Mathematics, Grade 12	April 2014

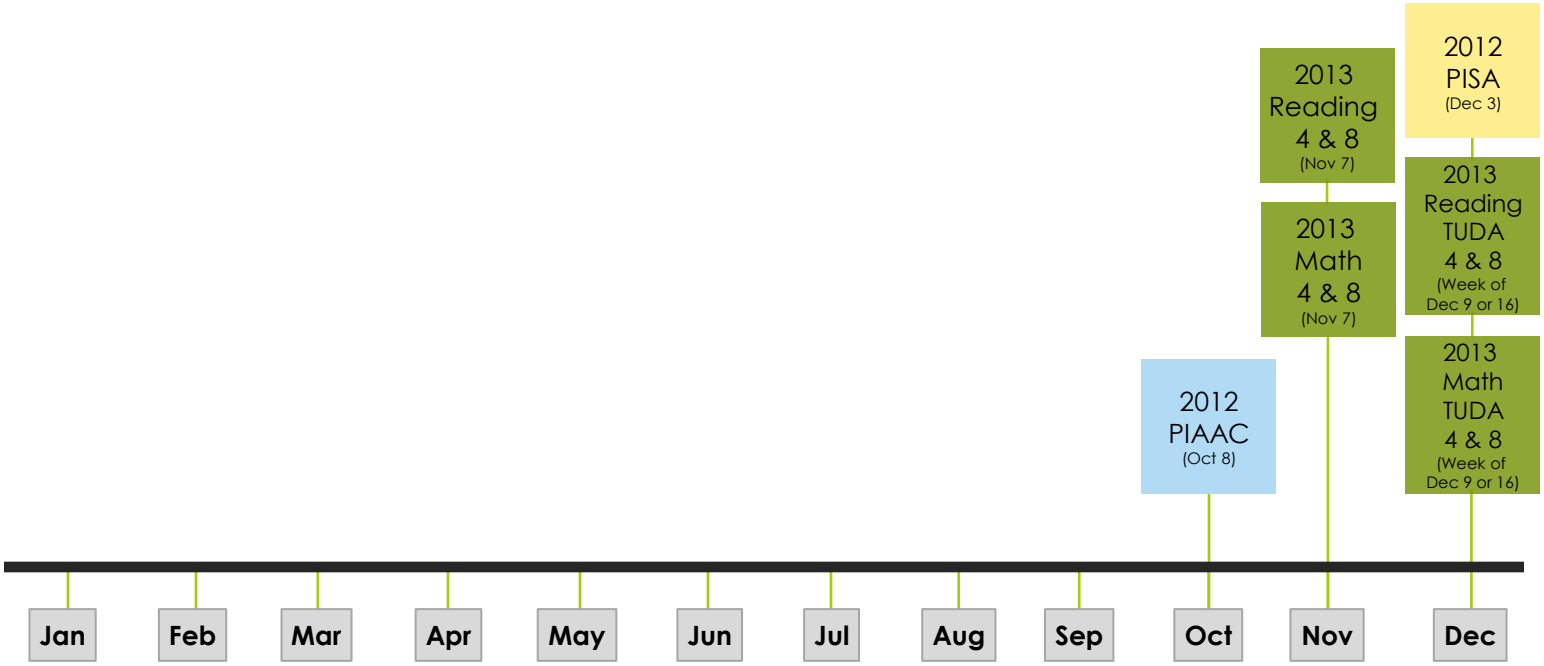
Other NAEP Reports

Grade 8 Black Male Students: Through the lens of the National Assessment of Educational Progress	February 2014
Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade- (Technical Report)	June 2014

Other Related Reports from NCES

Performance of U.S. 15-Year Old Students in Mathematics, Science, and Reading Literacy in An International Context First Look at PISA 2012	December 2013

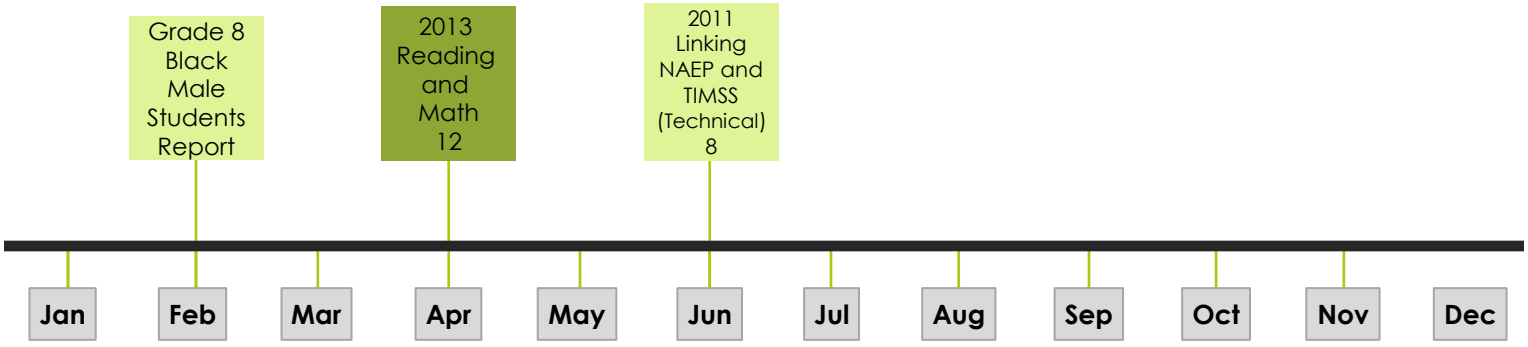
2013 NCES Assessment Data Release Timeline



LEGEND

- NAEP Report Cards
- PIAAC
- PISA

2014 NCES Assessment Data Release Timeline



LEGEND

- NAEP Report Cards
- NAEP Studies

Releases in 2013

- 2012 Program for the International Assessment of Adult Competencies (PIAAC)
- 2013 Reading Report Card: Grades 4 and 8
- 2013 Mathematics Report Card: Grades 4 and 8
- 2012 Programme for International Student Assessment (PISA)
- 2013 Reading Report Card: Trial Urban Districts (TUDA): Grades 4 and 8
- 2013 Mathematics Report Card: Trial Urban Districts (TUDA): Grades 4 and 8

Releases in 2014

- Grade 8 Black Male Students: Through the Lens of the National Assessment of Educational Progress
- 2013 Reading and Mathematics Report Card: Grade 12
- Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade (Technical Report)

Assessment Data Collection Schedule 2014

- U.S. History: Grade 8
- Civics: Grade 8
- Geography: Grade 8
- Technology and Engineering Literacy: Grade 8

Education Summit for Parent Leaders

Date: January 13, 2014

Location: Crystal Gateway Marriott Hotel, Arlington, VA

Summit Planning Group: Rebecca Gagnon, Terry Mazany, Tonya Miles, Eileen Weiser

Summit Workshop Advisors: Shannon Garrison, Doris Hicks, Dale Nowlin

Summit Workshop Consultants: Amanda Avallone, Kathi King (former Board members)

Summit Workshop Facilitators: Lou Fabrizio, Rebecca Gagnon, Shannon Garrison, Doris Hicks, Andrew Ho, Brent Houston, Hector Ibarra, Father Joseph O'Keefe, Terry Mazany, Tonya Miles, Dale Nowlin, and former Board members Robin Hall, Kim Kozbial-Hess, and Sister Mary Frances Taymans

Summit Goal: Convey the urgency of improving student achievement in the United States for all children and the urgency of reducing achievement gaps between student subgroups.

The Summit will enable attendees to use NAEP data and resources to ask the right questions of their education leaders about the status of student achievement and gaps in achievement locally, and to promote productive conversations about what is being done to improve achievement and close achievement gaps.

General Description: The day-long event has two major components. The morning plenary will consist of individual speakers and panel presentations from a range of perspectives on the need to raise student achievement and close achievement gaps. In the afternoon, hands-on workshops will familiarize participants with NAEP data and resources and illustrate their potential in framing questions relevant to educational improvement.

Audience: 150 parent leaders concerned about education from across the nation, with special attention to the 21 districts participating in the NAEP Trial Urban District Assessment Program. Sessions will be available across the nation via live-streaming internet video and/or live TV and radio coverage, if feasible.

Confirmed Speakers: Kati Haycock (the Education Trust); Marc Morial (National Urban League): Janet Murguia (National Council of La Raza); Steve Murdock (Rice University); Charles Payne (University of Chicago); Otha Thornton (National PTA)

Invited Speakers: First Lady Michelle Obama; Secretary of Education Arne Duncan.