

Appendix G

NAEP 2016 Sample Design Memo



An Employee-Owned
Research Corporation

1600 Research Boulevard
Rockville, MD 20850-3129
tel: 301-251-1500
fax: 301-294-2040
www.westat.com

Date: February 25, 2015

Memo: 2016-3.1PSU/1.1S

To: Bill Ward, NCES
Amy Dresher, ETS
Ed Kulick, ETS
Nancy Caldwell
Debby Vivari
Rick Rogers
William Wall
Leslie Wallace
Erin Wilson
Chris Averett
Kavemuii Murangi

David Freund, ETS
Nicole Beaulieu, ETS
Connie Smith, Pearson
Dianne Walsh
Lauren Byrne
Dwight Brock
Rob Dymowski
John Burke
Joel Wakesberg
Lloyd Hicks
Sabrina Zhang

From: Leslie Wallace

Reviewer: Dave Hubble

Subject: Sample Design for 2016 NAEP – Overview

I. Introduction

For 2016, the sample design involves several components, all of which are national assessments of one kind or another.

1. Operational assessment in Music at grade 8;
2. Operational assessment in Visual Arts at grade 8;
3. Technology-based Assessment (TBA) Pilot in Mathematics at grades 4 and 8;
4. TBA Pilot in Reading at grades 4 and 8;
5. TBA Pilot in Writing at grades 8 and 12;

In addition, the following special studies will be conducted:

6. Puerto Rico TBA Proof-of-Concept Study at grades 4 and 8
7. Oral Reading Fluency Study at grade 4

A small subset of the students selected for the grade 4 TBA Pilot will be identified and asked to complete an additional 10-minute block for the Oral Reading Fluency Study. No additional sample is necessary for this study.

The target sample sizes of assessed students for the remaining components are shown in Table 1 (which also shows an estimate of the required number of participating schools). All of these assessments are to take place in the typical NAEP testing window of late January to early March 2016.

Table 1. 2016 NAEP Sample Sizes (Public and Private)

	Session*	Public school students	Private school students	Total students
Grade 4				
Mathematics (P)	0401	13,500	1,500	15,000
Reading (P)		12,825	1,425	14,250
Puerto Rico (S)	0402	400	0	400
Total		26,725	2,925	29,650
Schools		540	220	760
Grade 8				
Mathematics (P)	0801	13,500	1,500	15,000
Reading (P)		8,325	925	9,250
Writing (P)		4,050	450	4,500
Puerto Rico (S)	0802	400	0	400
Music (O)	0803	3,600	400	4,000
Visual Arts (O)	0804	3,600	400	4,000
Total		33,475	3,675	37,150
Schools		725	235	960
Grade 12				
Writing (P)	1201	4,050	450	4,500
Total		4,050	450	4,500
Schools		215	25	240
GRAND TOTAL				
Schools		1,480	480	1,960

* The actual session codes to be used are unknown at this time.
(O) = Operational, (P) = Pilot, and (S) = Special Study

II. Assessment Types

From a sampling and operations point of view, many types of assessment sessions can be distinguished. The detailed target counts of assessed students are provided in Table 1. Below is a summary of major points.

1. The Music assessment at grade 8. This assessment will be conducted in the Arts PSUs. This assessment must be conducted in a different physical session from the Visual Arts assessment (see immediately below), but the two assessments may be in the same schools. The session has a target of 4,000 assessed students.
2. The Visual Arts assessment at grade 8. This assessment will be conducted in the Arts PSUs. This assessment must be conducted in a different physical session from the Music

assessment, but the two assessments may be in the same schools. The session has a target of 4,000 assessed students.

3. The TBA Pilot assessments for grades 4, 8, and 12 will be computer delivered. Because of the different delivery method, these assessments must be in separate sessions than the Arts assessments. In fact, out of concern for overburdening schools, an additional set of PSUs, with minimum overlap with the Arts PSUs, will be used for conducting the TBA Pilot assessments. For grades 4 and 8, in which two or three subjects will be assessed for the TBA Pilot, these subjects will be spiraled together. For grade 4, the spiral will include mathematics and reading. For grade 8, the spiral will include mathematics, reading, and writing. For grade 12, the TBA Pilot assessment will be writing only. For grade 4, the TBA Pilot session has a target of 29,250 assessed students (15,000 mathematics and 14,250 reading). For grade 8, the TBA Pilot session has a target of 28,750 assessed students (15,000 mathematics, 9,250 reading, and 4,500 writing). For grade 12, the TBA Pilot writing session has a target of 4,500 assessed students.
4. The Puerto Rico TBA Proof-of-Concept Study at grades 4 and 8. This is a TBA mathematics assessment. The session has a target of 400 assessed students for each of grades 4 and 8.

III. Primary Sampling Units Selection and Overlap Control

As the TBA Pilot assessments are national, with a total original sample size of assessed students of about 62,500, for reasons of operational efficiency in conducting the assessments a sample of Primary Sampling Units (PSUs) was selected, and all sampled schools were drawn from within the sampled PSUs. With a smaller sample size of about 8,000 assessed students for the Arts assessment at grade 8, a separate sample of PSUs was selected with the largest PSUs being in common to both PSU samples.

The PSUs were created from aggregates of counties. Data on counties were obtained from the 2010 Census, and the definitions of Metropolitan Statistical Areas (MeSAs) used were the December 2009 Office of Management and Budget (OMB) definitions. Each Metropolitan Statistical Area (MeSA) constitutes a PSU, except that MeSAs that cross Census region boundaries were split into their individual regional components.

Non-metropolitan PSUs were formed by aggregating counties into geographic units of sufficient minimum size to provide enough schools to constitute a workload of about 1% of the total sample. These PSUs were made of contiguous counties where possible, and almost contiguous counties (separated by MeSA counties) otherwise. Each PSU falls within a single state.

This process generated a frame of approximately 1,000 PSUs. The PSUs were stratified, using characteristics aggregated from county-level characteristics, found by analysis to be related to NAEP achievement in past assessments. A sample of 67 PSUs was selected for the TBA Pilot sample. The 29 largest MeSAs were selected with certainty, and the remaining sample was a stratified probability proportional to size (PPS) sample, where the size measure was a function of the number of children as given in the most recent population estimates prepared by the U.S. Census Bureau. For the TBA Pilot sample, 76 such strata were formed and paired and a single PSU was selected from one stratum in each of the 38 pairs for a total of 67 PSUs. For the Arts sample, the same certainty PSUs were selected. However, in order to ensure no overlap in sampled schools a single PSU was selected from each of the strata in the 38 pairs not selected for the TBA Pilot sample for a total of 67 PSUs.

Note that the PSUs for the NAEP 2016 TBA Pilot samples were selected in such a way as to minimize overlap with the NAEP 2015 sample PSUs, the NAEP 2014 paper-and-pencil sample PSUs, and the NAEP 2013 sample PSUs. This was done to reduce the chance that a school was selected for the 2016 TBA Pilot and one (or more) of the 2015, 2014 paper-and-pencil or 2013 assessments. Similarly, the PSUs for the NAEP 2016 Arts samples were selected in such a way to minimize overlap with the NAEP 2015 sample PSUs and the NAEP 2014 computer-based sample PSUs. This was done to reduce the chance that a school was selected for the 2016 Arts assessment and one (or more) of the 2015 assessments or the 2014 computer-based assessments.

IV. Stratification and Oversampling

As in the recent past, the plan was to draw separate public and private school samples. This approach has proven to be useful, in that, selecting the samples separately has three advantages: 1) it permits the timing of sample selection to vary between public and private schools, should this prove necessary; 2) it allows us to readily assume different response and eligibility rates for public schools and private schools; and 3) it makes it easier to use different sort variables for public schools and private schools. It also allows for the possibility of a late change of mind concerning the sample sizes that differ between public and private schools. Note that the Puerto Rico TBA Proof-of-Concept Study did not include a private school component.

Explicit stratification took place at the PSU level. For schools within PSUs, stratification gains were achieved by sorting the school file prior to systematic selection. As in past national samples, the expectation was that, within the set of certainty MeSA PSUs within a census region, PSU would not necessarily be the highest level sort variable. Thus, type of location was used as the primary sort variable. Consider for example the large MeSAs in the Midwest region. The design was aimed primarily at getting the correct balance of city, suburban, town, and rural schools, as a priority over getting exactly a proportional representation from each MeSA (Chicago, Detroit, Minneapolis), although of course it should be possible to get a high degree of control over both of these characteristics. The sort of the schools used other variables beyond the type of location variable, such as a race/ethnicity percentage variable. The exact set of variables used in sorting the schools prior to sampling is specified in the particular sampling specification memos.

In addition, we implemented oversampling of high-minority schools within the public schools. That is, as used in past national assessments, a public school with over 15 percent Black and Hispanic combined enrollment was given twice the chance of selection of a public school of the same size with a lower percentage of these two groups. This approach is effective in increasing substantially the sample sizes of Black and Hispanic students, without inducing undesirably large design effects on the sample, either overall or for particular subgroups. This oversampling was performed for both the TBA Pilot and Arts samples. Beyond this, we will also implement the oversampling of Black and Hispanic students at the student level in schools not being oversampled at the schools level, that is, schools with less than 15 percent Black and Hispanic students, for the Arts assessment only.

The updated preliminary 2013/14 CCD and the updated 2013/14 PSS school files were approved for use by NCES. They served as the public and private school frames for the 2016 NAEP.

V. New Schools

To compensate for the fact that the CCD file used to create the NAEP public school sampling frames is out of date at the time of frame construction, we supplemented the samples for the Arts

assessment with a sample of new public schools for grade 8. The frame used to select the sample was based on the NAEP 2015 new schools frame for grade 8. New school samples were not developed for the private school samples.

The new school samples were drawn using a two-stage design. At the first stage, conducted for NAEP 2015 and not repeated for NAEP 2016, a national sample of school districts was selected from the 2015 Operational and TEL sample PSUs. The sampled districts were asked to review lists of their respective schools and identify new schools. Frames of new schools were constructed from these updates. At the second stage, conducted for NAEP 2016, the NAEP 2015 grade 8 new school frame was first restricted to schools located in the Arts sample PSUs. Then this restricted set of schools was compared to the NAEP 2016 public school frame for the Arts sample (that was based on the 2012-2013 CCD) to see which schools should still be considered new for NAEP 2016. Once this frame had been constructed, new schools were drawn with probability proportional to size using the same sample rates as their corresponding original (Arts) school samples.

Note that the student and school sample sizes in Table 1 do not reflect these new school samples. However, some schools from the original sample proved to be closed or otherwise ineligible, and the new school procedure essentially compensates for the sample losses from these sources, as well as improving coverage of the population.

VI. Within-PSU Overlap Control with Other Samples

In keeping with the efforts at the PSU level to reduce potential overlap between the TBA Pilot and Arts samples, methods were employed to reduce overlap during sample school selection within the PSUs that contain more than one sample. With this approach it was possible to avoid any school overlap among the different school samples at a given grade. Schools were sometimes selected to participate at more than one grade.

The Keyfitz method was used to compute conditional probabilities to reduce the overlap between the samples within grade. That is, in the 29 certainty PSUs that overlapped between TBA Pilot and Arts, the conditional probabilities of selection for the TBA Pilot schools were based on the Arts schools sampling outcomes. Specifically, this was done to reduce overlap between Arts grade 8 sample schools and the TBA Pilot grade 8 sample schools.

VII. Substitute Samples

Substitute samples were selected for each of the 2016 samples in the following order: Arts grade 8, TBA Pilot grade 12, TBA Pilot grade 8, and TBA Pilot grade 4. This ordering of samples and grades is necessary since no school can be selected as a substitute more than once. It is more critical for operational samples to precede the pilot samples and higher grades to precede lower grades due to having fewer schools available to serve as substitutes at the higher grades. This was done separately for both public and private schools. The general steps for selecting substitutes were to put the school frames in their original sampling sort order, and take the 'nearest neighbor' of each original sampled school, excluding schools selected for any of the NAEP 2016 samples, schools already selected to serve as a substitute school, and schools which cross PSU or state boundaries, as potential substitutes.

The nearest neighbor is the school adjacent (immediately preceding or succeeding) the original school in the sorted frame with the closer estimated grade enrollment value. If estimated grade enrollment of both potential substitute schools differed from the original school by the exact same amount, the

selection procedure randomly chose one of the schools. If neither the preceding or succeeding school was eligible to be a substitute, then the sampled school was not assigned a substitute.

In addition, the few sampled private schools whose school affiliation was unknown did not get substitutes nor could such private schools not in sample serve as substitute schools. Also, new schools did not get substitute schools nor serve as substitutes.

VIII. Student Sampling

Student sample sizes within each school are determined as the combined result of several factors:

1. We wish to take all students in relatively small schools.
2. We wish to avoid the situation where all but a few students (e.g., more than 90%, but fewer than 100%) are tested.
3. We do not wish to have a sample that is too clustered for any one assessment subject.
4. We do not wish to have many physical sessions that contain only a very small number of students, as this is inefficient.
5. We do not wish to overburden the schools with unduly large student samples.
6. For the TBAs, we can use up to 25 tablets in a school at one time.

The plans below reflect the design that results from considering each of these factors and balancing them.

Arts: Grade 8 Schools

We will select all students, up to 40. In public schools with more than 40 students, student sampling depends on the percent of students that are Black and Hispanic. Specifically, in public schools with less than 15% Black and Hispanic students, we will select 35 plus an oversample of up to 5 additional Black and Hispanic students. Complementarily, in those public schools with more than 15% Black and Hispanic students (schools oversampled as described in Section IV), we will select 35 students. In private schools we will select all students up to 40, and sample 35 students in larger schools.

There are 16 session types: eight Music and eight Visual Arts. Half the students will be assigned to Music and half to Visual Arts. Minimum session size is 12 within schools with 12 or more students. In order to achieve the minimum session size, every school will not be assigned both session types. Session type assignment is detailed in Table 2.

To simplify the operational complexities, the number of permutations is substantially reduced by pairing each Music session type with one Visual Arts session type to create eight session type pairs. Assignment to these session type pairs will be done at the school level. Schools with 12 or more students will be assigned one session type pair. Smaller schools will be assigned a single session type of either Music or Visual Arts.

Table 2. Arts grade 8 school session type allocations and proportions

Enrollment size	Grade 8	
	1 to 23	24 and higher
Probability of being assigned Music and Visual Arts	0	1
Proportion of sample students assigned to Music (in schools with both session types)	NA	1/2
Probability of being assigned Music only	1/2	0
Probability of being assigned Visual Arts only	1/2	0

TBA Pilot: Grade 4 and 8 Schools

We will select all students, up to 50. In schools with more than 50 students we will select 50. All students will be assigned to the TBA Pilot.

TBA Pilot: Grade 12 Schools

We will select all students, up to 25. In schools with more than 25 students we will select 25. All students will be assigned to the TBA Pilot.

Puerto Rico TBA Proof-of-Concept Study: Grade 4 and 8 Schools

We will select all students, up to 25. In schools with more than 25 students we will select 25. All students will be assigned to the TBA session for their grade.

IX. Weighting Requirements

Arts Samples

The operational Arts samples will have a single set of weights for each subject (Music and Visual Arts at grade 8) applied to reflect probabilities of selection, school and student nonresponse, any trimming, and the random assignment to the particular subject. There will be a separate replication scheme by grade and public/private.

TBA Pilot Samples

We will not weight the students in the TBA Pilot samples at grades 4, 8, and 12. However, preliminary weights will be available for these pilot samples.

Puerto Rico Samples

We will not weight the students in the Puerto Rico TBA Proof-of-Concept Study at grades 4 and 8. However, preliminary weights will be available for these samples.

Appendix H
Sample Data Security Agreement
(used in e-filing process)



UNITED STATES DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION STATISTICS

Data Security Agreement for the
2015 National Assessment of Educational Progress (NAEP)

This document establishes a data security agreement between the _____ **State Department of Education**, and the **National Center for Education Statistics** of the U.S. Department of Education for the 2015 NAEP assessment program.

NAEP is a congressionally mandated project of the U.S. Department of Education. P.L. 107-279, Title III, directs the Commissioner for Education Statistics to conduct a National Assessment of Educational Progress. The law requires the NCES Commissioner for Education Statistics to conduct a national and state assessment in mathematics and reading in grades four and eight at least once every 2 years. At grade 12, NAEP assesses mathematics and reading at regularly scheduled intervals. Additional national assessments in grades 4, 8, and 12 take place at regularly scheduled intervals in other subjects such as writing, science, history, geography, civics, economics, foreign languages, and arts. This agreement applies to data collection activities under the NAEP program including operational, and pilot and special studies.

P.L. 107-110, as amended by P.L. 107-279 authorizes NAEP to include, “whenever feasible, information collected, cross-tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency.” To fulfill this statutory requirement, in addition to cognitive questions, NAEP administers background questionnaires that provide information for reporting categories and that collect non-cognitive data on students, their family background, teachers, and schools.

NCES understands that any improper disclosure or unauthorized use of these materials may violate Federal statutes, including but not limited to the Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. 1232g) as well as applicable state statutes.

By accepting this agreement, NCES acknowledges that student records and related information constitute confidential materials and commits to protect and safeguard these data according to NAEP data security procedures, as applicable, described in **NCES Data Confidentiality Procedures Summary** (Attachment A) and incorporated herein.

ACCEPTED AND AGREED TO:

State authorized agent Date

Title

Address

NCES authorized agent Date

Title

Address