NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
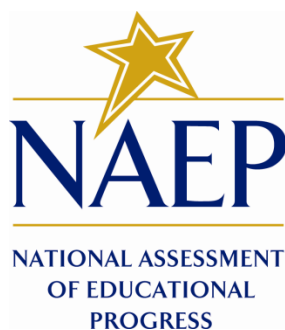
APPENDIX C
EXAMPLE OF SAMPLE DESIGN DOCUMENT
(2009 ASSESSMENT)

SYSTEM CLEARANCE PROPOSAL

NAEP SURVEYS

FOR THE YEARS 2011-2013

October 26, 2009

Memo

Date: June 27, 2008

Memo: 2009-2.3A/1.3B/1.3G/1.3D/1.2E

To: Holly Spurlock, NCES   Michael Lapp, ETS
Clyde Reese, ETS   David Freund, ETS
Connie Smith, Pearson   Dianne Walsh
Lauren Byrne   Debby Vivari
Rob Dymowski   William Wall
John Burke   Lou Rizzo
Susan Fuss   David Ferraro
Chris Averett   Kavemuii Murangi

From: Lloyd Hicks and Keith Rust

Subject: Sample Design for 2009 NAEP

Changes since previous version:

1)   Some details about the High School Transcript Study sample are provided;
2)   The session types for the ICT assessments have been amended;
3)   The student sampling specifications for the grade 12 beta and epsilon samples have been revised, to give a more efficient administration;
4)   Details of the student sampling plans for the grade 12 delta schools are included, having been omitted inadvertently in the previous version.

I.   Introduction

For 2009, the NAEP assessment involves several components:

■   National assessments in reading, math, and science at grades 4, 8 and 12;

■   State-by-state and Urban District (TUDA) assessments in reading, math, and science for public schools at grades 4 and 8;

■   State-by-state assessments in reading and math for public schools at grade 12, in eleven states;

■   National assessment probes in Science Hands-On Tasks (HOT) and Science Interactive Computer Testing (ICT) at grades 4, 8, and 12;

■   Pilot tests in reading and math at grades 4 and 8 and in civics, US history, and geography at grades 4, 8, and 12.

Below is a summary list of the features of the 2009 sample design.

1.   As in recent NAEP studies, Urban District Assessment (TUDA) samples will form part of the corresponding state samples, and the state samples will form part of the national sample. The same ten participants as in 2007 will be involved, plus an additional seven new districts. They include Los Angeles, San Diego, Atlanta, Chicago, Boston, New York City, Charlotte, Cleveland, Austin, and Houston as the continuing districts, and Fresno, Miami-Dade, Jefferson County, KY, Baltimore, Detroit, Philadelphia, and Milwaukee.

2.   A major change from 2007 is that there will be two public school samples, and two private school samples at each grade. Small PSU-based samples, known as the beta samples (for public schools) and the epsilon samples (for private schools), will be used for science HOT and ICT, and all pilot tests. The other samples (the alpha sample for grades 4 and 8 public, the gamma sample for grade 12, delta for private schools at grades 4, 8, and 12) will be used for the operational assessments in reading, mathematics, and science, including trend/bridging samples for reading and mathematics.

3.   Unlike 2005 and 2007, there will be no large state oversampling for the reading, math, and science assessments at grades 4 and 8, nor any charter school oversampling.

4.   There is to be no special study of charter schools, and no oversampling of charter schools.

5.   At grades 4 and 8, all jurisdictions but Alaska, Kansas, Nebraska, Vermont, and the District of Columbia have signed on for science. All jurisdictions doing science will receive a 1:1:1 RS spiral, and the five that did not will receive a 9:9:1 RN spiral to ensure that these states are sufficiently represented in the national science sample. A 1:1:1 RS spiral means science is spiraled in at the same rate as reading and math. The 9:9:1 RN spiral means that for every 9 math and 9 reading booklets there will be 1 science booklet.

6.   At grades 4 and 8, all BIE schools sampled for the operational assessments will receive a 9:9:1 RN spiral.

7.   At grades 4 and 8, private schools sampled for the operational assessments will receive a 1:1:1 RS spiral (the RN spiral will not be used for private schools).

8.   All seventeen TUDA districts will do science in grade 4 and 8 since all their associated states have signed on for science. Thus for the operational assessments they will receive the RS spiral.

9.   At grade 12, there will be two different spirals for the operational reading, mathematics, and science assessments. One (denoted RS) will be used for providing national samples. The other (RM) will be used for providing supplemental samples for reading and mathematics in public schools in the eleven states participating in the state-by-state assessments. The eleven states participating at the state level for reading and

mathematics are Arkansas, Connecticut, Florida, Idaho, Illinois, Iowa, Massachusetts, New Hampshire, New Jersey, South Dakota, and West Virginia.

10. There will be no samples in territories, other than for Puerto Rico in math at grades 4 and 8. Schools samples are to be selected for Puerto Rico, although it is very likely that they will not be utilized for any assessments.

11. As in 2007, there will be larger samples of BIE schools in the grades 4 and 8 operational assessments than in earlier years. All BIE schools and students will be included in the sample. This is designed to provide detailed national results for American Indian and Alaskan Native (AIAN) students in reading and mathematics as part of the National Indian Education Study (NIES). Because science will be spiraled with math and reading, the science sample will be spread across almost all BIE schools at grades 4 and 8.

12. In a handful of states (Washington, Oregon, Arizona, North Carolina, Minnesota, and Utah) the public school sample at grades 4 and 8 will be increased somewhat, so as to give publishable results for AIAN students. This will affect school sampling only. There will be no special student sampling procedures for this purpose. This will be achieved by increasing, by an appropriate factor, the measures of size of schools that are in specially designated strata with a relatively high proportion of AIAN students.

13. As in 2007, the Department of Defense Schools will be reported as a single jurisdiction (DoDEA), instead of the two components of domestic (DDESS) and overseas (DoDDS). However, for design purposes, we will still sample and weight these as two separate entities.

14. As in 2007, at grades 4 and 8, private schools will be oversampled in sufficient numbers so as to be able to report Catholic and non-Catholic schools separately. There will be no special oversampling as implemented in 2002, 2003, and 2005 for grades 4 and 8. There will be no oversampling of private schools at grade 12.

15. There will be a High School Transcript Study (HSTS) conducted. This will be similar to that conducted in 2005 (there was no HSTS in 2007). The grade 12 schools and students from the mathematics and science (but not ICT and HOT) will be included in the HSTS, with the exception that school subsampling will be carried out in those states that are participating in the grade 12 state NAEP pilot.

The sample sizes of assessed students for these various components are shown in Table 1 (which also shows the approximate numbers of participating schools). Note that the sample size for 4th and 8th grade public schools for science reflects the appropriate Science sample from each of the five "nonparticipating" states and BIE schools in order to ensure nationally representative samples.

Table 1.   Target sample sizes of assessed students, and expected number of participating schools, for 2009 NAEP

	Spiral	Jurisdictions		Students		
	Spiral Indic.	States (incl. DC, BIE, DoDEA)	Urban districts	Public school students (with NIES)	Private school students	Total
Grade 4						
nat'l/state math	RS, RN	53	17	170,600	3,000	173,600
nat'l/state reading	RS, RN	53	17	170,600	3,000	173,600
nat'l/state science	RS, RN	47	17	156,000	3,000	159,000
Total - alpha	**2**			**497,200**		
Total - delta	**1 (RS)**				**9,000**	
Maximum students per school				90	90	
Average assessed students per school				54	24	
Total schools - alpha, delta				**9,207**	**375**	

Table 1.   Target sample sizes of assessed students, and expected numbers of participating schools for 2009 NAEP (cont'd)

	Spiral	Jurisdictions		Students		
	Spiral Indic.	States (incl. DC, BIE, DoDEA)	Urban districts	Public school students (with NIES)	Private school students	Total
Grade 4						
civics pilot test	PI			900	100	1,000
math pilot test	PI			2,700	300	3,000
reading pilot test	PI			4,050	450	4,500
National science ICT	CE, CS			1,800	200	2,000
National science HOT	SH			1,800	200	2,000
US history pilot test	PI			900	100	1,000
geography pilot	PI			900	100	1,000
Total beta, epsilon	**4**			**13,050**	**1,450**	
Maximum students per school				54	54	
Average assessed students per school				36	20	
Total schools – beta, epsilon				*363*	*73*	
math – PR	PR	1		3,000		3,000
Total alpha PR	**1**			**3,000**		
Maximum students per school				30		
Average assessed students per school				27		
Total public schools – alpha PR				*111*		
Total Number of Schools – grade 4				*9,681*	*448*	*10,129*
Total Number of Students – grade 4				**513,250**	**10,450**	**523,700**
Grade 8						
nat'l/state math	RS, RN	53	17	170,600	3,000	173,600
nat'l/state reading	RS, RN	53	17	170,600	3,000	173,600
nat'l/state science	RS, RN	47	17	156,000	3,000	159,000
Total – alpha	**2**			**497,200**		
Total – delta	**1 (RS)**				**9,000**	
Maximum students per school				90	90	
Average assessed students per school				70	25	
Total schools – alpha, delta				*7,103*	*360*	
civics pilot test	PI			900	100	1,000
math pilot test	PI			2,700	300	3,000
reading pilot test	PI			2,700	300	3,000
Nat'l science ICT	CE, CS			1,800	200	2,000
Nat'l science HOT	SH			1,800	200	2,000
US history pilot test	PI			1,350	150	1,500
geography pilot	PI			1,350	150	1,500
Total beta, epsilon	**4**			**12,600**	**1,400**	
Maximum students per school				52	52	
Average assessed students per school				40	24	
Total schools – beta, epsilon				*315*	*58*	

Table 1. Target sample sizes of assessed students, and expected numbers of participating schools for 2009 NAEP (cont'd)

	Spiral	Jurisdictions		Students		
	Spiral Indic.	States (incl. DC, BIE, DoDEA)	Urban districts	Public school students (with NIES)	Private school students	Total
Grade 8						
math - PR	PR	1		3,000		3000
Total alpha PR	**1**			**3,000**		
Maximum students per school				30		
Average assessed students per school				27		
Total public schools - alpha PR				*111*		
Total Number of Schools - grade 8				*7,529*	*418*	*7,947*
Total Number of Students - grade 8				512,800	10,400	523,200
Grade 12						
national math	RS			8,100	900	9,000
national reading	RS			9,900	1,100	11,000
national science	RS			9,900	1,100	11,000
reading trend (mixed)	RS			5,400	600	6,000
reading trend (old)	RS			6,300	700	7,000
math trend (mixed)	RS			5,400	600	6,000
math trend to 2005 (old)	RS			6,300	700	7,000
state math	RM	11		23,700		23,700
state reading	RM	11		23,700		23,700
Total - gamma	**2**			**98,700**		
Total - delta	**1 (RS)**				**5,700**	
Maximum students per school				80-150	150	
Average assessed students per school				70	30	
Total schools - gamma, delta				*1,410*	*190*	
reading pilot test				0	0	0
math pilot test				0	0	0
civics pilot test	PI			900	100	1,000
Nat'l science ICT	CE, CS			1,800	200	2,000
Nat'l science HOT	SH			1,800	200	2,000
US history pilot test	PI			900	100	1,000
geography pilot	PI			900	100	1,000
Total beta, epsilon	**4**			**6,300**	**700**	
Maximum students per school				52	52	
Average assessed students per school				35	25	
Total schools - beta, epsilon				*180*	*28*	
Total Number of Schools - grade 12				*1,590*	*218*	*1,808*
Total Number of Students - grade 12				105,000	6,400	111,400
GRAND TOTAL SCHOOLS				*18,800*	*1,084*	*19,884*
GRAND TOTAL STUDENTS				1,131,050	27,250	1,158,300

II. Assessment Types

In 2009, there will a total of eight different assessment types, or "spirals". At any one grade six or seven of these will be used. These different assessment types are summarized in Table 2. Session IDs contain 6 digits, traditionally. The first two digits identify the assessment "type" (subjects and type of spiral in a general way). Grade is contained in the second pair of digits, and the session sequential number (within schools) in the last 2 digits.

Note that the reading and mathematics assessments at each grade actually involve three subcomponents: new (corresponding to the new frameworks in each subject), old, corresponding to the most recent past assessments in these subjects (2007 for grades 4 and 8, 2005 for grade 12), and braided, using a combination of new and old material.

Table 2.   NAEP 2009 assessment types and IDs

ID	Type	Subjects	Grades	Schools	Comments
RM	Operational	Reading (new), math (new)	12	Public, in 11 participating states	For state-by-state reading and math results
RS	Operational	Reading, math, science	4, 8, 12	Public & private	1:1:1 reading, math, Science spiral at grades 4 and 8, for states signed on for Science, and private schools. At grade 12 the relative proportions are as needed for national samples, and will be used in all samples.
RN	Operational	Reading, math, science	4, 8	Public & BIE	9:9:1 reading, math, science spiral: For states not signed on for science, and BIE schools.
CE	Probe	Science ICT, Extended prompt first	4, 8, 12	Public & private	National sample
CS	Probe	Science ICT, Short prompt first	4, 8, 12	Public & private	National sample
SH	Probe	Science HOT	4, 8, 12	Public & private	National sample
PI	Pilot	Reading (4, 8), math (4, 8), civics, US history, geography	4, 8, 12	Public & private	National sample
PR	Puerto Rico	Some form of mathematics assessment	4 & 8	Public	Booklets will not be used in any other jurisdiction.

III. Sample Types and Sizes

In similar fashion to 2005 (but somewhat extended), we will identify five different types of school samples: Alpha, Beta, Gamma, Delta, and Epsilon). These distinguish sets of schools that will be conducting distinct portions of the assessment.

1.   Alpha Samples at Grades 4 and 8

These are public school samples for grades 4 and 8. They will be used for the operational state-by-state assessments in reading, math, and science, and contribute to the national samples for these subjects as well. There will be alpha samples for each state, DC, DoDEA, BIE, and Puerto Rico.

The details of the target student sample sizes for the alpha samples are as follows:

  A.   At each grade, the target student sample size for a state depends on whether or not it signed on for science. For each state that signed on for science, the target student sample size is 9,450: 3,150 each for math, reading, and science. In these states the RS session type will be used. For the four states not signed on for science, plus DC and BIE, the target student sample sizes are 6,650: 3,150 each for math and reading and 350 for science. In these jurisdictions the RN session type will be used.

  B.   In DC, DoDEA, BIE, and several small states, all students will be included (those states in which there are fewer students than are required in the above sample sizes, or just a few more than that).

  C.   There will be samples for seventeen TUDA districts. For the five largest (New York City, Los Angeles, Chicago, Miami-Dade, and Houston), the student target sample sizes are three-quarters the size of a state (7,088). For the other twelve districts the student target sample sizes are one-half the size of a state (4,725). All TUDA assessments will use the RS session type.

  D.   Note that above there is a conflict between sample size requirements at the state level, and the TUDA district level. This will be resolved as in previous years: the districts will have the target samples indicated in C. For the states that contain one or more of these districts, the target sample size indicated in A will be used to determine a school sampling rate for the state, which will be applied to the balance of the state outside the TUDA district(s).

  E.   In Puerto Rico, the target sample size is 3,150 per grade. Only mathematics will be assessed in Puerto Rico (PR session type).

As in past state-by-state assessments, schools with fewer than 20 students in the grade in question will be sampled at a moderately lower rate than other schools (at least half, and often higher, depending upon the size of the school). This is in implicit recognition of the greater cost and burden associated with surveying these schools.

As mentioned above, the NAEP 2009 design includes an oversample of high proportion American Indian schools (as part of the NIES design). These schools will be sampled at higher rates than the other schools. The NIES oversample will take place in Arizona, Minnesota, North Carolina, Oregon, Utah, and Washington. Schools with relatively large percentages of American Indian students will be separately stratified, as explained below, and oversampled by factors ranging from 2

to 6 based on state and grade. Table 3 below shows the thresholds used to define the NIES oversampling strata along with their corresponding oversampling factors.

Table 3.   Percent American Indian thresholds and oversampling factors for the NIES school oversample by state and grade

	Grade 4		Grade 8	
State	Percent American Indian thresholds	Oversampling factor	Percent American Indian thresholds	Oversampling factor
Arizona	50	3	50	2
Utah	5	5	5	5
Minnesota	10	5	10	4
North Carolina	15	6	10	6
Oregon	10	6	15	4
Washington	10	6	15	6

Table 4 shows the counts of schools to be selected in the alpha samples, along with the school and student frame counts, and final target student sample sizes by state and TUDA districts for grades 4 and 8. The school sample sizes reflect the undersampling of very small schools. However, they do not reflect the impact of the oversampling of high percentage American Indian schools in the six states listed in Table 3. The table also identifies the jurisdictions where we take all schools and where we take all students.

Table 4.   Grade 4 and 8 school and student frame counts, expected school sample sizes, and initial target student sample sizes for the 2009 state-by-state and TUDA district assessments (alpha samples)

	Grade 4					Grade 8				
Jurisdiction	Schools in frame	Schools in sample	Students in frame	Overall target student sample size		Schools in frame	Schools in sample	Students in frame	Overall target student sample size	
Alabama	755	138	58,253	9,450		480	120	59,066	9,450	
Alaska	359	200	9,660	6,650		281	136	10,254	6,650	
Arizona	1,110	137	80,785	9,450		713	130	79,822	9,450	
Arkansas	514	148	35,818	9,450		329	125	36,824	9,450	
Bureau Of Indian Education	140	140	3,839	3,839	**	120	120	3,561	3,561	**
California	5,698	316	478,659	24,526		2,550	235	476,256	24,644	
Colorado	967	156	58,127	9,450		466	124	58,659	9,450	
Connecticut	598	138	43,521	9,450		261	112	43,976	9,450	
Delaware	107	107	9,353	9,353	**	55	55	9,751	9,751	**
District Of Columbia	126	126	5,174	5,174	**	54	54	4,503	4,503	**
DoDEA Schools	131	131	8,239	8,239	**	67	67	5,742	5,742	**
Florida	1,968	184	203,335	15,262		1,066	167	202,242	15,316	
Georgia	1,174	173	121,798	13,332		484	125	123,356	12,839	
Hawaii	196	137	14,147	9,450		69	69	13,051	9,450	*
Idaho	349	162	19,940	9,450		189	110	20,248	9,450	

Table 4. Grade 4 and 8 school and student frame counts, expected school sample sizes, and initial target student sample sizes for the 2009 state-by-state and TUDA district assessments (alpha samples) (cont'd)

	Grade 4				Grade 8			
Jurisdiction	Schools in frame	Schools in sample	Students in frame	Overall target student sample size		Schools in frame	Schools in sample	Students in frame	Overall target student sample size	
Illinois	2,300	226	156,987	14,558		1,543	209	160,233	14,633	
Indiana	1,123	141	79,531	9,450		483	111	81,376	9,450	
Iowa	686	187	34,520	9,450		397	136	36,854	9,450	
Kansas	736	149	34,063	6,650		414	126	34,911	6,650	
Kentucky	747	193	49,113	12,828		398	143	49,880	12,914	
Louisiana	747	148	51,437	9,450		489	131	50,570	9,450	
Maine	360	217	14,494	9,450		220	136	15,710	9,450	
Maryland	864	203	62,185	13,144		320	147	65,742	13,242	
Massachusetts	1,021	213	71,651	12,942		472	139	75,007	13,444	
Michigan	1,974	213	123,765	13,396		1,055	169	131,165	13,508	
Minnesota	931	152	59,705	9,450		670	137	63,660	9,450	
Mississippi	443	134	38,179	9,450		285	114	39,291	9,450	
Missouri	1,128	163	67,206	9,450		673	133	71,518	9,450	
Montana	413	288	10,797	9,450		280	190	11,755	9,450	
Nebraska	705	196	21,193	6,650		405	139	22,112	6,650	
Nevada	339	119	32,723	9,450		144	86	33,157	9,450	
New Hampshire	266	164	15,249	9,450		136	97	16,301	9,450	
New Jersey	1,349	139	101,057	9,450		717	117	102,106	9,450	
New Mexico	417	162	24,512	9,450		207	112	25,659	9,450	
New York	2,343	177	198,642	13,294		1,253	161	209,701	13,497	
North Carolina	1,346	174	109,359	13,322		680	141	110,404	13,369	
North Dakota	276	276	7,140	7,140	**	197	197	7,830	7,830	**
Ohio	2,026	226	134,367	13,603		1,101	196	141,653	13,885	
Oklahoma	898	181	46,592	9,450		597	151	46,908	9,450	
Oregon	741	166	40,687	9,450		371	128	41,332	9,450	
Pennsylvania	1,799	194	130,786	13,180		867	159	142,456	13,309	
Puerto Rico	1,022	110	44,476	3,150		411	105	41,856	3,150	
Rhode Island	186	160	11,109	9,450		60	60	11,998	9,450	*
South Carolina	597	125	52,356	9,450		277	109	54,236	9,450	
South Dakota	334	334	8,999	8,999	**	269	269	9,624	9,624	**
Tennessee	975	139	72,426	9,450		561	120	71,722	9,450	
Texas	3,971	271	342,664	20,625		2,060	181	333,813	21,230	
Utah	524	130	39,959	9,450		205	108	38,108	9,450	
Vermont	227	227	6,807	6,807	**	126	126	7,047	7,047	**
Virginia	1,120	129	89,984	9,450		378	107	95,321	9,450	
Washington	1,200	147	75,870	9,450		604	123	79,640	9,450	
West Virginia	440	203	20,525	9,450		200	117	21,555	9,450	
Wisconsin	1,128	249	60,347	13,164		618	179	64,652	13,231	
Wyoming	194	194	6,234	6,234	**	99	99	6,503	6,503	**

Table 4.   Grade 4 and 8 school and student frame counts, expected school sample sizes, and initial target student sample sizes for the 2009 state-by-state and TUDA district assessments (alpha samples) (cont'd)

Urban districts (TUDA)	Grade 4					Grade 8				
	Schools in frame	Schools in sample	Students in frame	Overall target student sample size		Schools in frame	Schools in sample	Students in frame	Overall target student sample size	
Atlanta	62	62	4,209	4,209	**	21	21	3,670	3,670	**
Austin	76	61	6,360	4,725		21	21	5,247	5,247	**
Baltimore City	123	86	6,790	4,725		61	49	6,499	4,725	
Boston	78	78	4,022	4,022	**	34	34	4,569	4,569	**
Charlotte	90	55	9,856	4,725		35	35	9,398	4,725	*
Chicago	462	109	32,784	7,088		444	108	32,186	7,088	
Cleveland	81	81	4,468	4,468	**	83	83	4,753	4,753	**
Detroit	148	70	10,184	4,725		83	55	9,227	4,725	
Fresno	67	57	6,219	4,725		25	25	5,965	4,725	*
Houston	193	93	16,701	7,088		58	47	14,268	7,088	
Jefferson County	93	63	6,998	4,725		36	36	6,657	4,725	*
Los Angeles	483	85	57,291	7,088		121	73	51,742	7,088	
Miami	227	83	27,406	7,088		98	65	26,074	7,088	
Milwaukee	125	91	6,437	4,725		89	64	6,481	4,725	
New York City	699	89	68,125	7,088		373	83	67,430	7,088	
Philadelphia	176	66	13,752	4,725		123	59	13,061	4,725	
San Diego	135	67	10,340	4,725		47	35	9,824	4,725	

1) Counts for states include counts for their constituent TUDA districts.
2) Target student sample sizes reflect sample sizes prior to attrition due to exclusion, ineligibility, and nonresponse.
3) * identifies jurisdictions where all schools for the given grade are included in the NAEP sample.
4) ** identifies jurisdictions where all students for the given grade are included in the NAEP sample.

Stratification

In the six states where the NIES oversampling is taking place (Arizona, Minnesota, North Carolina, Oregon, Utah, and Washington), we will first create a separate stratum of schools with relatively large percentages of American Indians (those above the thresholds given in table 3). These oversampling strata will not be further broken down. The remaining schools will be further stratified as described below.

Each state and grade will be stratified separately, but using a common approach in all cases. TUDA districts will be separated from their state, and each part stratified separately. The first level of stratification will be based on urban-centered type of location. This variable has 12 levels (some of which may not be present in a given state or TUDA district), and these will be collapsed so that each of the resulting location categories contains at least 9 percent of the student population. Within each of the resulting location categories, schools will be assigned a minority enrollment status. This is

based on the two race/ethnic groups that are the second and third most prevalent within the location category. If these groups are both low in percentage terms, no minority classification will be used. Otherwise three (or occasionally four) equal-sized groups (generally high, medium, and low minority) will be formed based on the distribution across schools of the two minority groups.

Finally, within the resulting location and minority group classes (of which there are likely to be from three to fifteen, depending upon the jurisdiction), schools will be sorted by a measure derived from school level results from the most recent available state achievement tests at the relevant grade. In general, mathematics test results will be used, but where these are not available, reading results will be used. In the few states that do not have math or reading tests at grades 4 and 8 (or where we are unable to match the results to the NAEP school frame), instead of achievement data, schools will be sorted using a measure of socio-economic status. This is the median household income of the 5-digit ZIP Code area where the school is located, based on 2000 Population Census data. Schools in the American Indian oversampling strata (in the six states having such strata) will be sorted by percentage of American Indian enrollment.

Once the schools are sorted by location class, minority enrollment class, and achievement data (or household income or AIAN enrollment data), a systematic sample of schools will be selected using a random start. Schools will be sampled with probability proportional to size. The exact details of this process are described in the individual sampling specification memos.

2.   Beta Sample

The beta samples at grades 4, 8, and 12, are public school samples. These samples will be used to conduct the IC, SH, and PI assessments in public schools.

The first stage of sampling for the beta samples is the selection of a sample of geographic primary sampling units (PSUs). These will be selected using the same design as for the 2008 Long Term Trend assessment, but minimizing the overlap of with the PSUs in that assessment. A total of 62 PSUs will be selected, representing the U.S. (but not including Puerto Rico, or DODEA schools that are located outside the 50 states and D.C.).

This PSU component is needed because of the operational complexities of administering the science ICT assessment. A select group of staff will be trained to administer ICT. The HOT (SH) and pilot test (PI) samples included in the beta sample for efficiency. This makes the school samples more efficient, and reduces complications in spiraling.

The school stratification of the beta sample within PSUs will be by type of location and median household income.

3.   Gamma Sample

This is the public school sample at grade 12, used for the reading, mathematics, and science assessments. It will consist of a single sample of schools but with a mixture of two assessment types. Schools in the eleven state-by-state participants will each receive a mixture of these two assessment

types, which can be administered with a common administration, whereby students assigned to either of these assessment types can be assessed in the same room, under common conditions.

As in past assessments, modest oversampling of Black and Hispanic students will be undertaken in this sample. This will be carried out at the school level. Each school with more than 15 percent Black and Hispanic students will be given twice the selection probability of a low minority school of comparable size. This means that while about 53 percent of the student population (including over 90 percent of the Black and Hispanic students) are in "high" minority schools, about 70 percent of the sample students will come from these schools. This oversampling will not be applied in the eleven states participating in the state-by-state assessment.

In the eleven states participating in the state-by-state trial/pilot, each school in the sample will be assigned a combination of the RM and RS session types. In the balance of the US the RS session type will be used.

Stratification

Unlike the other samples, the Gamma sample will have an explicit stratification at the highest level. The schools will be stratified by state, for the eleven state-by-state participants, and the balance of the US, giving 12 explicit strata. In each of the eleven states, stratification will be carried out in the same way as for the alpha samples. The last level sorting variable will be median household income, as there are no achievement data available at the school level for grade 12.

For the explicit stratum containing the balance of the US, the highest level of sorting will be by Census Division. This gives 8 implicit strata (the balance of New England is too small to be an effective stratum, and is combined with the balance of the Mid-Atlantic Division). The next stratifier in the hierarchy is type of location, which has twelve categories. Many of the type of location strata nested within Census divisions will be collapsed with neighboring type of location cells (this will occur if the expected school sample size within the cell is less than 4.0). These geographic strata will be subdivided using a dichotomous high minority status category. Schools are in the high minority stratum if they had more than 10 minority students and greater than 15 percent minority students (minority defined as Black or Hispanic). Otherwise the school will be put in a low minority stratum. If the expected sample size within these strata is less than 8.0, they will be left as is. If the expected sample size is greater than 8.0, then the high-or-low minority strata will be subdivided into up to four substrata (two for expected sample size up to 12.0, three for expected sample size up to 16.0, and four for expected sample size greater than 16.0). For the low minority strata, the subdivision will be by state or groups of contiguous states. For the high minority strata, the subdivision will be by minority percentage. Within these substrata, the schools are to be sorted by school type (public, BIE, DoDEA) and median household income from the 2000 Census (using a serpentine sort within the school type substrata).

4.   Delta Samples

These are the private school samples at grades 4, 8, and 12, for conduction the operational assessments in reading, math, and science. Schools in the three grade-specific samples will be assigned the RS session type.

Stratification

The private schools are to be explicitly stratified by private school type (Catholic/non-Catholic). Within each private school type, stratification will be by Census region (4 categories), type of location (12 categories), and enrollment size. In general, where there are few or no schools in a given stratum, categories will be collapsed together always preserving the Catholic/non-Catholic stratification.

5.   Epsilon Samples

These samples are analogous to the beta samples, but for private schools. The same PSUs will be used as for the beta samples, and the IC, SH, and PI session types will be conducted. The school stratification within PSUs will be based on Catholic/non-Catholic status, type of location, and enrollment size.

IV.   New Schools

To compensate for the fact that files used to create the NAEP school sampling frames are at least two years out of date at the time of frame construction, we will supplement the Alpha, Gamma, and Delta samples with new school samples at each grade.

The new school samples will be drawn using a two-stage design. At the first stage, a minimum of ten school districts (in states with at least ten districts) will be selected from each state for public schools, and ten Catholic dioceses will be selected nationally for the private schools. The sampled districts and dioceses will be asked to review lists of their respective schools and identify new schools. Frames of new schools will be constructed from these updates, and new schools will be drawn with probability proportional to size using the same sample rates as their corresponding original school samples.

The school sample sizes in the above tables do not reflect new school samples.

V.   Substitute Samples

Substitute samples will be selected for each of the Beta, Gamma, Delta, and Epsilon samples. The substitute school for each original will be the next "available" school on the sorted sampling frame, with the following exceptions:

1.   Schools selected for any NAEP samples will not be used as substitutes.

2.   Private schools whose school affiliation is unknown will not be used as substitutes. Also, unknown affiliated private schools in the original samples will not get substitutes.

3.   Schools can only be substitutes for one and only one sample. (If a school is selected as a substitute school for grade 12, for example, it can not be used as a substitute for either grade 4 or grade 8.)

4.   A public school substitute will always be in the same state as its original school.

VI.   Student Sampling

Students within the sampled schools will be selected with equal probability. The student sampling parameters vary by sample type (Alpha, Beta, Gamma, Delta, and Epsilon), assessment type, and grade, as described below.

Sample Alpha, Grade 4 Schools (Except Puerto Rico), and Sample Delta, Grade 4

For alpha schools in states signed on for Science, and delta schools:

1.   All students, up to 102, will be selected.

2.   If the school has between 103 and 120 students (inclusive), the school will be asked if it wishes to have all students selected, or a sample of 90. If the school elects to have all students selected, all students will be selected. If the school asks for a sample of 90 students, a systematic sample of 90 students will be selected.

3.   If the school has more than 120 students, a systematic sample of 90 students will be selected.

4.   In some schools, the school may be assigned more than one 'hit' in sampling. In these schools we will select a sample of size 90 times the number of hits, taking all students if this target is greater than or equal to 90/102 of the total grade 4 enrollment.

5.   All students will be assigned to assessment type RS.

For alpha schools in states not signed on for Science:

1.   All students, up to 68, will be selected.

2.   If the school has between 68 and 120 students (inclusive), the school will be asked if it wishes to have all students selected, or a sample of 63. If the school elects to have all students selected, all students will be selected. If the school asks for a sample of 63 students, a systematic sample of 63 students will be selected.

3.   If the school has more than 120 students, a systematic sample of 63 students will be selected.

4.   In some schools in small jurisdictions, the school may be assigned more than one 'hit' in sampling. In these schools we will select a sample of size 63 times the number of hits, taking all students if this target is greater than or equal to 63/68 of the total grade 4 enrollment.

5.   All students will be assigned to assessment type RN.

Sample Alpha, Grade 8 Schools (Except Puerto Rico), and Sample Delta, Grade 8

For alpha schools in states signed on for Science, and delta schools:

1.   All students, up to 102, will be selected.

2.   If the school has more than 102 students, a systematic sample of 90 students will be selected with no oversampling.

3.   In some schools, the school may be assigned more than one 'hit' in sampling. In these schools we will select a sample of size 90 times the number of hits, taking all students if this target is greater than or equal to 90/102 of the total grade 8 enrollment.

4.   All students will be assigned to assessment type RS.

For alpha schools in states not signed on for Science:

1.   All students, up to 68, will be selected.

2.   If the school has more than 68 students, a systematic sample of 63 students will be selected with no oversampling.

3.   In some schools in small jurisdictions, the school may be assigned more than one 'hit' in sampling. In these schools we will select a sample of size 63 times the number of hits, taking all students if this target is greater than or equal to 63/68 of the total grade 8 enrollment.

4.   All students will be assigned to assessment type RN.

Sample Alpha, Puerto Rico Grades 4 and 8

1.   All students, up to 34, will be selected.

2.   If the school has more than 34 students, a systematic sample of 30 students will be selected with no oversampling.

3.   All students will be assigned to assessment type PR.

Samples Beta and Epsilon, Grades 4, 8, and 12

The session types involves in these samples are the two ICT session types (CE and CS), the HOT (SH), and the pilot tests (PI). When a school is assigned to do ICT, the session type of CE or CS will be assigned at the school level, with half of the ICT schools doing each type.

At grade 4, 2/29 of the students will be assigned to CE, 2/19 to CS, and 4/29 to HOT (SH), with 21/29 assigned to the pilot tests (PI). No school will be assigned more than one of CS, CE, or SH. Thus in a school of sufficient size there will be a sample of 54 students, with 14 or 15 students assigned to either CS, CE, or SH and the remainder to PI. In schools with fewer than 16 students, all students will be assigned to one of CS, CE, SH, or PI.
At grade 8, 1/14 of the students will be assigned to CE, 1/14 to CS, and 1/7 to HOT (SH), with 5/7 assigned to the pilot tests (PI). No school will be assigned both more than one of CS, CE, and SH. Thus in a school of sufficient size there will be a sample of 52 students, with 14 or 15 students assigned to either CE, CS, or SH and the remainder to PI. In schools with fewer than 16 students, all students will be assigned to one of CS, CE, SH, or PI.

At grade 12, 1/7 of the students will be assigned to CE, 1/7 to CS, and 2/7 to HOT (SH), with 3/7 assigned to the pilot tests (PI). Thus in a school of sufficient size there will be a sample of 52 students, with 14 or 15 students assigned to either CE or CS, 14 or 15 assigned to SH, and the remainder to PI. In schools with between 16 and 26 students inclusive, 4/7 of the students will do one of CE, CS, or SH, and 3/7 will do PI. One quarter of such schools will do CE, one quarter will do CS, and one half will do SH. In schools with fewer than 16 students, all students will be assigned to one of CE, CS, SH, or PI.

Samples Gamma and Delta, Grade 12

For schools in states not signed on for the state-by-state assessments in reading and math for gamma, and for all schools in delta:

1.   All students, up to 170, will be selected.

2.   If the school has more than 170 students, a systematic equal probability sample of 150 students will be selected.

3.   All students will be assigned to assessment type RS.

For gamma schools in Florida:

1.   All students, up to 150, will be selected.

2.   If the school has more than 150 students, a systematic equal probability sample of 125 students will be selected.

3.   Three-fifths of the students in each school will be assigned to assessment type RM, and 2/5 to RS.

For gamma schools in Illinois:

1. All students, up to 140, will be selected.

2. If the school has more than 140 students, a systematic equal probability sample of 120 students will be selected.

3. Five-eighths of the students in each school will be assigned to assessment type RM, and 3/8 to RS.

For gamma schools in New Jersey:

1. All students, up to 120, will be selected.

2. If the school has more than 120 students, a systematic equal probability sample of 100 students will be selected.

3. Three-quarters of the students in each school will be assigned to assessment type RM, and 1/4r to RS.

For gamma schools in Massachusetts:

1. All students, up to 110, will be selected.

2. If the school has more than 110 students, a systematic equal probability sample of 95 students will be selected.

3. 15/19 of the students in each school will be assigned to assessment type RM, and 4/19 to RS.

For gamma schools in Arkansas, Connecticut, and Iowa:

1. All students, up to 100, will be selected.

2. If the school has more than 100 students, a systematic equal probability sample of 85 students will be selected.

3. 15/17 of the students in each school will be assigned to assessment type RM, and 2/17 to RS.

For gamma schools in Idaho, New Hampshire, South Dakota, and West Virginia:

1. All students, up to 95, will be selected.

2.   If the school has more than 95 students, a systematic equal probability sample of 80 students will be selected.

3.   15/16 of the students in each school will be assigned to assessment type RM, and 1/16 to RS.

VIII.   Weighting Requirements

The Operational Samples

These samples will have a single set of weights for each subject (reading, math, and science at grades 4, 8 and 12) 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.

The exact nature of the weighting for the reading assessments at all three grades, and the mathematics assessment at grade 12, will depend upon the decisions as to which of the components of new, mixed, and old booklets will be used in reporting. These decisions will be informed by the actual assessment data. It is possible that multiple sets of weights will be required for some of these assessments. Thus it will most likely not be possible to produce weights for the reading assessments at grades 4 and 8 on a schedule that is consistent with six-month reporting.

The Probe Samples

There will also be a set of weights for the national Science HOT assessment, and an additional set for the Science ICT assessment, at each grade. These weights will render the data representative of the U.S.

The Pilot Test Samples

We will not weight the students in the pilot test studies. However, preliminary weights will be available for pilot test samples, if needed.

The NIES Samples

The NIES survey samples consist of two grade-specific samples, comprising students selected for each of the grade 4 and 8 operational samples. We will create one set of weights for each grade-specific sample. The NIES weights are designed for any aggregation of the NIES data, not involving NAEP achievement data. NIES analyses involving NAEP achievement data should use the appropriate NAEP operational weights.

The HSTS Samples

The students eligible for the HSTS sample will be those gamma and grade 12 delta sample students, from either the RS or RM sessions, assigned either a mathematics or a science booklet. However, the complete set of such students is too large for the resources available for the HSTS. Thus some subsampling will be required, as follows:

  a) In ten of the eleven states participating in the pilot grade 12 program, a subsample of the gamma sample schools will be selected. Florida is the exception; there all of the gamma sample schools will be retained for the HSTS. In the 10 states the subsampling will be such that the rate of sampling in the HSTS from these states will be the same as from the remaining states not participating in the pilot state program. Thus the HSTS sample will be a national sample, supplemented by a state sample in Florida.

  b) Depending upon available resources, students taking the 'old' mathematics assessment booklets may be dropped from the HSTS sample.

More details of the HSTS subsampling process, and the resulting school and student sample sizes, will be provided in other internal Westat 2009 sampling memos.