APPENDIX A

Sample Design

Design Overview

School Sample
- Target Population
- School Sampling Frame
- School Stratification
- School Sample Allocation
- School Sample Selection
- Replacement Schools

Teacher Sample
- Target Population
- Teacher Sampling Frame
- Teacher Stratification
- Teacher Sample Selection
- Selection of Classes

Weighting and Standard Errors
- School Weights
- Teacher Weights
- Imputation of Number of Classrooms
- Calculating Standard Errors

Sample Design

Design Overview

The sample design for the 2012 National Survey of Science and Mathematics Education is a national probability sample of schools and teachers in grades K–12 in the 50 states and the District of Columbia. The sample was designed to allow national estimates (totals and ratios of totals) of science and mathematics course offerings and enrollment, teacher background and preparation, textbook usage, instructional techniques, and availability and use of science and mathematics facilities and equipment. Every eligible school and teacher in the target population had a known, positive probability of being drawn into the sample.

The sample design involved clustering and stratification. The first stage units consisted of elementary and secondary schools. Science and mathematics teachers constituted the second stage units. From the science and mathematics classes taught by sample teachers, a sample of one class was selected for each teacher. The target sample sizes were 2,000 schools and 10,000 teachers selected within sample schools. These sample sizes are large enough to allow sub-domain estimates such as for particular regions or types of community.

The sampling frame for the school sample was constructed from constructed from the Common Core of Data (CCD) and Private School Survey (PSS) databases—programs of the U.S. Department of Education's National Center for Education Statistics—that include school name and address and information about other characteristics needed for stratification and sample selection. The sampling frame for the teacher sample was constructed from lists provided by sample schools identifying active teachers and the specific science and mathematics subjects they were teaching.

School Sample

This section describes the sample design features of the school sample. It is organized as follows:

- Target Population;
- School Sampling Frame;
- School Stratification;
- School Sample Allocation;
- School Sample Selection; and
- Replacement Schools.

Target Population
The target population for the school sample includes all regular public and private schools in the 50 states and the District of Columbia. Excluded from the target universe are vocational/ technical schools, schools offering alternative, special or adult education only, and preschool/kindergarten-only schools.

School Sampling Frame
The sampling frame for the school sample was constructed from the final 2008–09 CCD and 2007–08 PSS public use files. Educational institutions classified as public, private, and Catholic elementary and secondary schools were included. Excluded were Bureau of Indian Affairs schools, Department of Defense schools, schools in Puerto Rico and the territories, alternative schools (e.g., special education, vocational, early childhood centers), ungraded schools, and pre-kindergarten only schools. For all schools in the database, CCD/PSS included information on grade span by indicating the lowest and highest grade offered in the school.

School Stratification
Three primary sampling strata were defined for the school sample. The strata definitions are based on grade span as follows:

- Stratum 1: Schools with any grade 10, 11, or 12;
- Stratum 2: Schools not in stratum 1, but with no grades lower than 5; and
- Stratum 3: All other schools.

Within primary strata, schools were further stratified by Census region (Midwest, Northeast, South, West), school metro status (rural, suburban/town, urban), and school type (public, private), resulting in a total of 72 strata.

School Sample Allocation
The allocation of the total school sample (2,000 schools) among the three primary strata was based on the minimum sample size desired for each stratum and the desired sample sizes for teachers of advanced mathematics and physics/chemistry. The sample allocation was the following:

- Stratum 1: 1,040 schools;
- Stratum 2: 480 schools; and
- Stratum 3: 480 schools.

School Sample Selection
Prior to sampling, schools were sorted by the first three digits of zip code (ZIP3) and total number of teachers within secondary strata. A serpentine sort was employed to sort schools from smallest to largest within ZIP3, then largest to smallest within the next ZIP3.

Schools were sampled within strata using probability proportional to size (PPS) systematic sampling, with measure of size equal to the total number of FTE teachers (public schools) or the total number of teachers (private schools) in the school. Schools with measure of size less than the 20th percentile for their stratum were assigned the 20th percentile as a measure of size to avoid large weights. In 1.1 percent of the schools, the total number of teachers was imputed using the average pupil-teacher ratio for the cells formed by the cross-classification of stratum (1-72), school locale (see Table A-1 for definition), school type (public, Catholic, non-Catholic religious, other private) and the school's reported enrollment:

Total teachers = Total enrollment / average(pupil-teacher ratio).

Table A-1
Definition of School Locale Code, Based on School Address.

Locale Code	Definition
11	City, Large Territory inside an urbanized area and inside a principal city with pop >= 250,000
12	City, Mid-size Territory inside an urbanized area and inside a principal city with pop < 250,000 and >= 100,000
13	City, Small Territory inside an urbanized area and inside a principal city with a population < 100,000
21	Suburban, Large Territory outside a principal city and inside an urbanized area with pop >= 250,000
22	Suburban, Mid-Size Territory outside a principal city and inside an urbanized area with a pop < 250,000 and >= 100,000
23	Suburb, Small Territory outside a principal city and inside an urbanized area with a pop < 100,000
31	Town, Fringe Territory inside an urban cluster <= 10 miles from an urbanized area
32	Town, Distant Territory inside an urban cluster > 10 miles and <= 35 miles from an urbanized area
33	Town, Remote Territory inside an urban cluster > 35 miles from an urbanized area
41	Rural, Fringe Census-defined rural territory <= 5 miles from an urban area; also rural territory <= 2.5 miles from an urban cluster
42	Rural, Distant Census-defined rural territory > 5 miles and <= 25 miles from an urbanized area; also rural territory > 2.5 miles and < 10 miles from an urban cluster
43	Rural, Remote Census-defined rural territory > 25 miles from an urbanized area and > 10 miles from an urban cluster

Replacement Schools

Five replacement schools were designated for each sampled school in case of nonresponse for the originally sampled school. The fifth replacement school was intended for a pilot study and was not used in the main study. The five replacement schools were usually the two or three schools listed just before and after the sampled school on the frame, after sorting as described above. The replacement schools were ranked by similarity with the sampled school with respect to number of teachers and assigned an "order of use" number so that the closest matching school within the same stratum/ZIP3 would be used first.

Table A-2 shows the distribution of the sample by primary and secondary stratum.

Table A-2
Distribution of Sample, by Stratum

	Secondary Stratum			Primary Stratum			
	Region	**Status**	**Public/ Private**	**1 Grades 10–12**	**2 Grades 5–9**	**3 Other**	**All Grades**
1	Midwest	Urban	Public	47	20	25	92
2			Private	11	0	4	15
3		Suburban	Public	92	56	41	189
4			Private	11	0	6	17
5		Rural	Public	65	22	25	112
6			Private	4	0	2	6
7	Northeast	Urban	Public	40	20	23	83
8			Private	13	0	4	17
9		Suburban	Public	102	59	45	206
10			Private	18	0	6	24
11		Rural	Public	37	15	13	65
12			Private	4	0	1	5
13	South	Urban	Public	83	54	50	187
14			Private	27	0	5	32
15		Suburban	Public	135	89	76	300
16			Private	26	0	5	31
17		Rural	Public	125	58	53	236
18			Private	14	9	1	15
19	West	Urban	Public	62	31	33	126
20			Private	13	0	4	17
21		Suburban	Public	72	40	44	152
22			Private	9	0	3	12
23		Rural	Public	72	12	14	32
24			Private	3	0	0	3
	TOTAL			**1,045**	**476**	**479**	**2,000**

Teacher Sample

The following section describes the sample design features of the teacher sample. It is organized as follows:

- Target Population;
- Teacher Sampling Frame;
- Teacher Stratification;
- Teacher Sample Selection; and
- Selection of Classes.

Target Population
The target population for the teacher sample consists of teachers in eligible schools (see School Sample, Target Population) who teach science and/or mathematics. Science includes biology, chemistry, physics, earth science, and other science.

Teacher Sampling Frame
The sampling frame for the teacher sample was constructed by requesting that coordinators in all sample schools provide a list of eligible teachers and identify the courses taught by each teacher. For schools in stratum 1, coordinators listed teachers in one of the following categories:

- High school physics or chemistry;
- Other science;
- High school calculus or advanced mathematics; and
- Other mathematics.

For strata 2 and 3, the categories listed were:

- Science and
- Mathematics

Teacher Stratification
Based on the course information provided for teachers on the school list, each teacher was assigned to one of the following five teacher strata:

- Physics/chemistry with or without other science, no mathematics;
- Advanced mathematics with or without other mathematics, no science;
- Other science only;
- Other mathematics only; and
- Any combination of mathematics and science.

Teacher Sample Selection
The goal was to sample about 10,000 teachers. Within each sampled school, seven teachers were sampled with probability proportional to a measure of size that was designed to oversample advanced math and physics/chemistry teachers at a rate of three. In schools with fewer than seven science/mathematics teachers, all teachers were selected. Prior to sampling, teachers were sorted by teacher stratum. The resulting sample sizes were:

- Primary Stratum 1: 5,561 teachers;
- Primary Stratum 2: 2,435 teachers; and
- Primary Stratum 3: 2,230 teachers.

The sampling fraction for teachers in teacher stratum *l* (l = 1 – 5) was computed as follows:

$$f_l = \frac{n_l}{N_l}$$

where:

f_l = Overall stratum sampling fraction in teacher stratum *l*

n_l = Number of teachers sampled in stratum *l*
N_l = Number of listed teachers in stratum *l*

For each of the three school primary strata, Table A-3 shows the number of teachers selected in the cooperating schools and the overall sampling fraction in each teacher stratum.

Table A-3
Teachers Selected in Each School Stratum

	Sample Size (n_l)	**Sampling Fraction (f_l)**
School Stratum 1: Grades 10–12	**5,561**	
1. Physics/chemistry with or without other science, no mathematics	1,593	0.5147
2. Advanced mathematics with or without other mathematics, no science	1,489	0.5129
3. Other science only	949	0.2216
4. Other mathematics only	1,164	0.2256
5. Any combination of science and mathematics	366	0.4164
School Stratum 2: Grades 5–9	**2,435**	
1. Physics/chemistry with or without other science, no mathematics	0	0
2. Advanced mathematics with or without other mathematics, no science	0	0
3. Other science only	1,013	0.4922
4. Other mathematics only	1,210	0.4852
5. Any combination of science and mathematics	212	0.4371
School Stratum 3: Other	**2,230**	
1. Physics/chemistry with or without other science, no mathematics	0	0
2. Advanced mathematics with or without other mathematics, no science	0	0
3. Other science only	113	0.3951
4. Other mathematics only	259	0.4060
5. Any combination of science and mathematics	1,858	0.3301

Selection of Classes

As part of the sampling process, teachers in sub-stratum five in each stratum were randomly assigned to receive either a science or a mathematics questionnaire. This step represented an additional stage of sampling since only half of the sample teachers in this stratum were assigned to report on science and the other half on mathematics. This one-in-two sub-sampling must be reflected in producing science- or mathematics-specific estimates.

Some of the items on the questionnaire apply to individual classes. Teachers with multiple science or mathematics classes each day were asked to report on only one of these classes. Teachers were asked to list all of their science and mathematics classes in order by class period. The web questionnaire used a pre-generated sampling table to make a selection from among the classes listed. The sampling table was randomly generated so that a random selection of classes would be achieved overall.

Weighting and Standard Errors

In surveys involving complex, multistage designs such as this national survey, weighting is necessary to reflect the differential probabilities of selection among sample units at each stage of selection. Weights were developed to produce unbiased estimates of the population of schools and teachers. Weighting is also used to adjust for different rates of participation in the survey by different types of schools and teachers. The final adjusted weights permit the respondents from the sample to represent the population of schools and teachers.

Variance computation must also take into account the survey design using a method such as jackknife or BRR replication, or Taylor series linearization. Statistical software packages that assume simple random sampling are not appropriate because they will underestimate the standard errors. To accommodate the sample design used in this study, a set of 75 jackknife (JK2) replicate weights was created for each full-sample school and teacher weight.[1]

Three school weights were developed corresponding to the School Coordinator Questionnaire, Science Program Questionnaire, and the Mathematics Program Questionnaire. Separate teacher and class weights were developed for the Science and Mathematics Teacher Questionnaires.

School Weights

Weights were developed to permit unbiased estimates for school and teacher characteristics. The base weight associated with a school is the reciprocal of the school's probability of selection and is calculated as follows:

$$W_{hi} = \frac{\sum_{i=1}^{N_h} MOS_{hi}}{n_h MOS_{hi}}$$

where:

MOS_{hi} = measure of size for school *i* in stratum *h*
N_h = total number of schools on the frame in stratum *h*
n_h = number of schools sampled in stratum *h*
h = 1, 2, …..72.

Replacement schools were used to substitute for non-cooperating schools, and for these the probability of selection of the originally sampled school was used to calculate the base weight. Of the 2,007 schools in the final sample, 749 were replacement schools and 7 were new schools, each formed by the merger of two schools on the frame after the sample was selected. The probability of selection for the new schools was calculated to take into account their increased

[1] Rust, K. F. and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research*, *5*(3), 283–310.

chance of selection. If the schools were from the same stratum, the probabilities of selection for the two schools that merged were summed. If they were from different strata, the probability of selection was calculated as 1 – (1-p(school 1))*(1-p(school 2)).

To adjust for different rates of participation in the survey by different types of schools, school nonresponse adjustments were developed and applied to the base weight.[2]

In some schools, the school coordinator questionnaire may not have been completed. In addition, the person designated to answer questions about the school science or mathematics program may have failed to participate. Accordingly, three distinct school non-response adjustments were developed:

- NRA1: To produce school estimates from the school coordinator questionnaire
- NRA2: To produce mathematics program level estimates
- NRA3: To produce science program level estimates

For non-response adjustment cell c, the general form of the NRA is given by:

$$NRA_c = \frac{\sum_{i \in \text{elig in c}} w_i}{\sum_{i \in \text{resp in c}} w_i}$$

where w_i is the base weight of the i[th] school in cell c. The numerator of the three adjustment factors is the same—all eligible schools. The denominator (respondents) for NR1 includes all schools that completed the school coordinator questionnaire; respondents for NR2 and NR3 include only schools that completed a program questionnaire in science and mathematics, respectively. As the replacement schools already compensate for non-response, the weights for these schools are included in the denominators of the adjustments.

Because nonresponse adjustment through weighting assumes that response patterns of non-respondents are similar to that of respondents, c corresponds to cells formed from school characteristics that were determined to be correlated with nonresponse. These characteristics were identified through a logistic regression model that predicted response propensity as a function of school characteristics. The characteristics identified by the model as correlated with response were school type (public, catholic, other private), primary stratum (grades 10–12, grades 5–9, other), high minority enrollment (> 25%), and metro status (urban, suburban, rural).

[2] For a discussion of nonresponse adjustments, see:

Kalton, G. and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology, 12*(1), 1–16.

Brick, J.M. and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*(215), (http://smm.sagepub.com/cgi/content/abstract/5/3/215)

The three school weights adjusted for non-response are given by:

$W_{1i,\,nr} = w_i * NR1_c$
$W_{2i,\,nr} = w_i * NR2_c$
$W_{3i,\,nr} = w_i * NR3_c$

where:

w_i = Base weight associated with school *i*
$NR1_c$ = Non-response adjustment factor for school coordinator questionnaire for schools in cell *c*
$NR2_c$ = Non-response adjustment factor for school mathematics programs in cell *c*
$NR3_c$ = Non-response adjustment factor for school science programs in cell *c*.

The non-response adjusted school weights were trimmed to the 99^{th} percentile of the weight distribution to reduce the effect of a few extremely large weights. These outlier weights arose from a few very small private schools that had a very small probability of selection. The weights that were not trimmed received a small adjustment so that the sum of the final school weights would equal the total of the school weights before trimming.

Teacher Weights

The teacher base weight is equal to the inverse of the overall probability of selection of the teacher, including the school's probability of selection. The teacher base weight was calculated as:

Teacher base weight = final school weight * (1/teacher probability of selection)

where the final school weight was adjusted for schools who refused to allow sampling of their teachers. (This was essentially the same set of schools that did not complete the school coordinator questionnaire.) Each teacher responded to either the science or mathematics questionnaire, but not both. For teachers sampled in the 5^{th} teacher stratum (both mathematics and science taught), the teacher probability of selection includes a factor of 2 to reflect the random assignment of these teachers to mathematics or science with a probability of 1/2.

The teacher base weight was adjusted separately for nonresponse to the mathematics and science questionnaires, as separate weights were planned for mathematics and science teachers. That is,

$$W_{ijk,\,nr} = \text{final school weight}_i * \text{teacher base weight}_{ij} * NRT_{jk}$$

where:

$W_{ijk,\,nr}$ = nonresponse-adjusted weight teacher *j* in school *i*, subject *k*,
NRT_{jjk} = nonresponse adjustment factor for teacher *j* in school *i*, subject *k*,
k = math or science.

NRT_{ijk} was calculated within adjustment cell c for each subject k as:

$$NRT_c = \frac{\sum_{\text{j} \in \text{elig in c}} w_{ij}}{\sum_{\text{j} \in \text{resp in c}} w_{\text{ij}}}$$

where w_{ij} is the base weight for teacher j in school i.

The nonresponse adjustment factor was calculated within adjustment cells formed using variables that were determined to be correlated with teacher nonresponse. These variables were identified using logistic regression models to predict response propensity to the mathematics and science teacher questionnaires as a function of school characteristics and teacher stratum. The variables identified by the model for both subjects were school level (grades 10–12, grades 5–9, other), school type (public, catholic, other private), high minority enrollment (>25%) and region (Northeast, Midwest, South, West). The unweighted response rate for both the mathematics and science questionnaires was 77 percent.

Because a small number of secondary teachers incorrectly identified themselves on the questionnaire as self-contained teachers, a second set of teacher weights was calculated for nonresponse to the class schedule item. The nonresponse adjustment factor was calculated within cells formed by variables correlated with nonresponse to this item, given the teacher was a respondent to the mathematics or science questionnaire. These variables were identified from a logistic regression model as school level (grades 10–12, grades 5–9, other), school type (public, catholic, other private), school size (small, medium, large) and teacher stratum. The unweighted response rate for this item was 94 percent, given the teacher was a respondent to the mathematics or science questionnaire.

The nonresponse-adjusted teacher weights were trimmed to a threshold of 4*average teacher weight to prevent extremely large weights, and the remaining teacher weights received a small adjustment factor to preserve the sum of the nonresponse-adjusted teacher weights prior to trimming. Five percent of the teacher weights were trimmed.

Imputation of Number of Classrooms

The number of classrooms taught was imputed when missing for teachers who responded to the mathematics or science questionnaires, including teachers who were deemed to have reported teaching only one self-contained classroom in error. The number of classrooms was imputed from another randomly selected teacher within the same teacher stratum and school, when possible, using the hot deck method.[3] If such a teacher could not be found, a teacher from the same teacher stratum within another school in the same school stratum, size, and minority class was selected. The number of classrooms was imputed for five percent of teacher respondents. Nearly two-thirds of the imputed values came from a teacher within the same school.

[3] Andridge, R. R. and Little, R. J. A. (2010). A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review*, *78*(1), 40–64.

Calculating Standard Errors

Estimates obtained from a sample of teachers will differ from the true population parameters because they are based on a randomly chosen subset of the population, rather than on a complete census of all mathematics and science teachers. This type of error is known as sampling error**.** The differences between the estimates and the true population values can also be caused by nonsampling error. Nonsampling errors can result from many causes, such as measurement error, nonresponse, sampling frame errors, and respondent error. The precision of an estimate is measured by the standard error (defined as the square root of the variance due to sampling). The calculation of the standard error must reflect the manner in which the sample was drawn. Otherwise, the standard errors can be misleading and result in incorrect confidence intervals and p-values in hypothesis testing. The study's sampling involved stratification, clustering, and unequal probabilities of selection, all of which must be reflected in the standard error calculations.

Replication methods such as the jackknife are commonly used to estimate variances for complex surveys involving multi-stage sampling. Replication methods work by dividing the sample into subsample replicates that mirror the design of the sample. A weight is calculated for each replicate using the same procedures as for the full-sample weight. This produces a set of replicate weights for each sampled school and teacher. To calculate the standard error of a survey estimate, the estimate is first calculated for each replicate using the replicate weight and the same form of estimator as for the full sample. The variation among the replicates is then used to estimate the variance for the full sample estimate, as given below in the formula for jackknife replicates formed with two variance units or pseudo-psus (primary sampling units) per stratum (JK2)[4]:

$$\mathrm{var}(\hat{\theta}) = \sum_{g=1}^{G} (\hat{\theta}_{(g)} - \hat{\theta})^2$$

where G is the total number of replicates $\hat{\theta}_{(g)}$ and is the estimate of $\hat{\theta}$ based on the observations included in the g[th] replicate.

For the current study, a set of 75 jackknife replicate weights was created for each school and teacher weight for calculating standard errors for school and teacher estimates. These may be used with packages that accommodate replication methods, such as WesVar, SUDAAN, Stata, or the survey procs in SAS v9.

[4] Rust, K. F. and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research*, *5*(3), 283–310.