Sampling and Weighting for 2018 NSSME+

Sampling

School Sampling Frame
School Stratification
Allocation of School Sample Size
Sample Selection of Schools
Replacement Schools
Target Population for Teacher Sampling
Teacher Sampling Frame
Teacher Stratification
Teacher Sample Selection
Selection of Science or Mathematics Classes

Weighting and Variances

School Weights
Teacher Weights
Calculating Standard Errors

Sampling and Weighting for 2018 NSSME+

Sampling

The 2018 NSSME+ used a stratified two-stage probability sample of science, mathematics, and computer science teachers in grades K–12 in the United States. At the first stage, 2,000 elementary and secondary schools were selected within strata with probability proportional to size (PPS). Although the final sampling plan projected 1,200 schools to participate in the survey, about 1,300 participated (65 percent response rate). At the second stage, approximately 10,000 science and mathematics teachers were sampled at predetermined rates to ensure a sufficient sample size for domain estimates, such as region or community type. Computer science teachers were sampled with certainty to allow for national estimates, as their prevalence in secondary schools is much lower than science and mathematics teachers. About 7,000 teachers were projected to complete the survey (70 percent response rate).

School Sampling Frame

The target population for the school sample includes all regular public and private schools in the 50 states and the District of Columbia. The school sampling frame was created from the final 2014–15 Common Core of Data (CCD) and the 2011–12 Private School Survey (PSS) public use file. The following types of school were excluded from the frame:

- Schools in Puerto Rico and the territories;
- Schools run by the Department of Defense;
- Schools run by the Bureau of Indian Education;
- Schools that are special education, vocational, technical, alternative, adult, career, virtual schools, or early childhood/child care centers;
- Schools that were closed or not yet open;
- Schools that are ungraded; and
- Schools that offer only Pre–K.

School Stratification

Schools on the frame were stratified by three primary strata using the CCD and PSS information on grade span: (1) school has any of grades 10–12, (2) school does not have any of grades 10–12 and has no grade lower than 5, and (3) all other schools. Within primary strata, schools were further stratified by Census region (Northeast, North Central, South, West), school metro status (urban, suburban/town, rural), and school type (public, private), resulting in a total of 72 strata.

Allocation of School Sample Size

The allocation of the 2,000 school sample size among the primary strata was based on the minimum sample size desired by stratum and the desired sample sizes for teachers of advanced mathematics and physics/chemistry. As in the 2012 National Survey of Science and Mathematics Education, 52 percent were allocated to primary stratum 1 and 24 percent were allocated to each primary stratum 2 and 3. Within primary strata, school sample sizes were secondary stratum. Sample sizes for each secondary stratum are displayed in Table A-1. The distribution of the sample across primary and secondary strata can be seen in Table A-2.

Table A-1
School Sample by Census Region, Metro Status, and School Type

REGION	SAMPLE SIZE	METRO STATUS	SAMPLE SIZE	SCHOOL TYPE	SAMPLE SIZE
Midwest	427	Urban	595	Public	1,770
Northeast	397	Suburban	995	Private	230
South	812	Rural	410		
West	364				
Total	**2,000**	**Total**	**2,000**	**Total**	**2,000**

Table A-2
Distribution of School Sample, by Stratum

	SECONDARY STRATUM			PRIMARY STRATUM			
	REGION	**METRO STATUS**	**PUBLIC/ PRIVATE**	**1 GRADE 10–12**	**2 GRADE 5–9**	**3 OTHER**	**ALL GRADES**
1	Midwest	Urban	Public	45	19	25	89
2			Private	12	0	5	17
3		Suburban	Public	92	61	45	198
4			Private	14	0	6	20
5		Rural	Public	54	19	22	95
6			Private	5	0	2	7
7	Northeast	Urban	Public	41	18	22	81
8			Private	18	1	4	23
9		Suburban	Public	100	61	44	205
10			Private	20	0	6	26
11		Rural	Public	30	12	12	54
12			Private	7	0	3	10
13	South	Urban	Public	89	58	53	200
14			Private	30	1	5	36
15		Suburban	Public	148	103	83	334
16			Private	26	0	6	32
17		Rural	Public	99	46	41	186
18			Private	18	0	3	21
19	West	Urban	Public	63	31	33	127
20			Private	15	0	5	20
21		Suburban	Public	76	42	42	160
22			Private	10	0	5	15
23		Rural	Public	22	8	9	39
24			Private	4	0	1	5
	TOTAL			**1,038**	**480**	**482**	**2,000**

Sample Selection of Schools

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 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 7.1 percent of the schools on the school frame, the total number of teachers was imputed using the average pupil-teacher ratio for the stratum (1–72) by school locale (see Table A-3 for definitions), by 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-3
Definition of School Locale Code, Based on School's 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

Four replacement schools were designated for each sampled school in case of nonresponse for the originally sampled school. The four replacement schools were usually the two or three schools listed just before and just 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.

Target Population for Teacher Sampling

The target population for the teacher sample consists of teachers in eligible schools (see School Sampling Frame section) who teach science and/or mathematics, or computer science.

Teacher Sampling Frame

The sampling frame for the teacher sample was constructed by requesting that principals in all sample schools appoint a study coordinator to provide a list of eligible teachers and identify the courses taught by each teacher. To assist the school in providing the information necessary to build the frame, an online form was provided to collect teaching categories depending on the school's primary stratum. For schools in primary stratum 1 the following categories were listed:

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

For primary 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 six teacher strata:

- Physics/chemistry with or without other science, no mathematics or computer science;
- Advanced mathematics with or without other mathematics, no science or computer science;
- Other science only;
- Other mathematics only;
- Any combination of mathematics and science, but no computer science; and
- Computer science regardless of other subjects taught.

Teacher Sample Selection

The goal was to sample about 10,000 teachers and get completed teacher questionnaires for 7,000 teachers. The target sample sizes were nine teachers per Grade 10–12 school, eight teachers per Grade 5–9 school, and seven teachers per Other school. If the number of teachers in the school was less than or equal to the target, all teachers were selected. All computer science teachers were selected with certainty in Grade 10–12 schools. For the remaining subjects, teachers were sampled with probability proportional to a measure of size that was designed to oversample advanced mathematics and physics/chemistry teachers at a rate of 3. Prior to sampling, teachers were sorted by teacher stratum. The resulting sample sizes were:

- Primary school stratum 1: 5,517 teachers;
- Primary school stratum 2: 2,356 teachers; and
- Primary school stratum 3: 2,066 teachers.

The sampling fraction for teachers in teacher stratum l (l = 1–6) 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

Table A-4 shows the number of teachers selected in the cooperating schools for each of the three primary school strata, and the overall sampling fraction in each teacher stratum. The sample sizes do not include 35 teachers who were sampled but later dropped because their school or district refused to participate after data collection began.

Table A-4
Teachers Selected in Each School Stratum

	SAMPLE SIZE (N_1)	SAMPLING FRACTION (F_1)
School Stratum 1: Grades 10–12	**5,517**	**0.4165**
1. Physics/chemistry with or without other science, no mathematics	1,428	0.5689
2. Advanced mathematics with or without other mathematics, no science	1,406	0.5871
3. Other science only	897	0.2752
4. Other mathematics only	1,060	0.2767
5. Any combination of science and mathematics	331	0.3899
6. Computer science	395	0.9850
School Stratum 2: Grades 5–9	**2,356**	**0.5672**
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	1,021	0.5688
4. Other mathematics only	1,217	0.5671
5. Any combination of science and mathematics	116	0.5498
6. Computer science	2	1.0000
School Stratum 3: Other	**2,066**	**0.3561**
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	118	0.3806
4. Other mathematics only	199	0.4243
5. Any combination of science and mathematics	1,749	0.3483

Selection of Science or Mathematics Classes

Sampled teachers were mailed invitations to complete an online questionnaire. 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 represented an additional stage of sampling since only half of the sampled 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 Variances

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 for school and teacher characteristics. 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.

Three school weights were developed corresponding to the School Coordinator Questionnaire, Science Program Questionnaire, and the Mathematics Program Questionnaire. A fourth school weight was developed for schools that completed teacher sampling and agreed to participate with the study, which was used in creating teacher weights. Three separate teacher weights were also developed for the Mathematics, Science, and Computer Science Teacher Questionnaires.

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.[24]

School Weights

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.

[24] Rust, K. and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research: Special Issue on the Analysis of Complex Surveys, 5,* 283–310.

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 2,008 schools in the final sample (including 9 newly merged schools discovered after sampling), 750 were replacement schools. The probability of selection for the new schools was calculated to take into account their increased 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))

because sampling was independent across strata.

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.[25]

Schools that did not allow teacher sampling were treated as nonresponding schools. In some schools, the School Coordinator Questionnaire was not completed. In addition, the person designated to answer questions about the school science or mathematics program may have failed to participate. Accordingly, four distinct school nonresponse adjustments were developed:

- NR1: To produce school estimates from the School Coordinator Questionnaire
- NR2: To produce mathematics program level estimates
- NR3: To produce science program level estimates
- NR4: To produce a school weight for calculating teacher weights

For nonresponse adjustment cell c, the general form of the nonresponse adjustment (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 or mathematics, respectively. The denominator for NR4 includes all schools that completed teacher sampling and agreed to cooperate. Since the replacement schools already compensate for nonresponse, the weights for these schools are included in the denominators of the adjustments.

Because nonresponse adjustment through weighting assumes that response patterns of nonrespondents are similar to that of respondents, c corresponds to cells formed from school

[25] 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)

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

characteristics that were determined to be correlated with nonresponse. These characteristics were identified through a tree classification program (SAS Proc HPSPLIT) that classified schools into cells ("leaves") defined by school characteristics, based on their response rates. The characteristics identified as correlated with response rates were school type (public, catholic, other private), high minority enrollment (> 25 percent), and metro status (urban, suburban, rural). Primary stratum (grades 10–12 , grades 5–9, other) was also used for public schools, since their larger numbers in the sample allowed four variables to be used to form nonresponse adjustment cells.

The four school weights adjusted for nonresponse are given by:

$$W_{1i,\,nr} = w_i * NR1_c$$
$$W_{2i,\,nr} = w_i * NR2_c$$
$$W_{3i,\,nr} = w_i * NR3_c$$
$$W_{4i,\,nr} = w_i * NR4_c$$

where:

w_i = Base weight associated with school *i*

$NR1_c$ = Nonresponse adjustment factor for School Coordinator Questionnaire for schools in cell *c*

$NR2_c$ = Nonresponse adjustment factor for Mathematics Program Questionnaire for schools in cell *c*

$NR3_c$ = Nonresponse adjustment factor for Science Program Questionnaires for schools in cell *c*

$NR4_c$ = Nonresponse adjustment factor for school teacher sampling in cell *c.*

The nonresponse 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 that refused to allow sampling of their teachers. Each teacher responded to only one of the mathematics, science, or computer science teacher questionnaires. For teachers sampled in the 5^{th} teacher stratum (both math and science taught), the teacher probability of selection includes a factor of 2 to reflect the random assignment of these teachers to math or science with a probability of 1/2.

The teacher base weight was adjusted separately for nonresponse to the mathematics, science, and computer science teacher questionnaires, because separate weights were planned for mathematics, science, and computer 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_{ijk} = nonresponse adjustment factor for teacher j in school i, subject k,
k = mathematics, science, or computer science.

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

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

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 a classification tree program (SAS Proc HPSPLIT) that classified teachers into cells defined by school characteristics based on their response status to the math, science, and computer science questionnaires. The variables identified by the program as correlated with teacher response rates were school level (grades 10–12, grades 5–9, other), school type (public, catholic, other private), high minority enrollment (>25 percent), metro status (urban, suburban, rural) and region (Northeast, Midwest, South, West). The unweighted response rate for both the mathematics and science questionnaires was 78 percent; for the computer science questionnaire the unweighted response rate was 79 percent.

The nonresponse-adjusted teacher weights were trimmed to a threshold of 5*average teacher weight to prevent extremely large weights from having undue influence on the estimates and variances, and the remaining teacher weights received a small adjustment factor to preserve the sum of the nonresponse-adjusted teacher weights prior to trimming. The percentage of responding teacher weights that were trimmed was 3.6 percent for mathematics teachers, 3.4 percent for science teachers, and 1.4 percent for computer science teachers.

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, science, and computer 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 process 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)[26]:

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

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 SAS, Stata, R, SUDAAN or WesVar.

[26] Rust, K. and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research: Special Issue on the Analysis of Complex Surveys, 5,* 283–310.