AIR®
AMERICAN INSTITUTES FOR RESEARCH®

NATIONAL SURVEY OF
PUBLIC EDUCATION'S
RESPONSE TO COVID-19

METHODOLOGY BRIEF

Preliminary Technical Supplement

Authors: Cong Ye | Kirk Wolter (NORC) | Mike Garet | Jordan Rickles | Dean Resnick (NORC)
Samantha Neiman | Sarah Hodgman

OCTOBER 2020

Introduction

In an effort to better understand how public and charter school governing bodies responded to the COVID-19 pandemic in the 2019–20 school year, the American Institutes for Research (AIR) and our partner NORC at the University of Chicago conducted the National Survey of Public Education's Response to COVID-19. This web survey of public school districts focused on a nationally representative sample of 2,536 public school districts across the United States. The data collection period was May 20, 2020, through September 1, 2020.

This methodology brief describes the sampling and weighting methods for this study, along with the final response rate.

Sample Design

The goal of the sampling strategy was to select a representative sample to enable unbiased estimates with reasonable precision within projected domains of analysis. Specifically, the sample was designed to enable state-level estimates for public school districts for 12 states: California, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, North Carolina, Oregon, Texas, Virginia, and Washington. For districts outside of these states, AIR strived to have adequate (for desired levels of precision) samples by U.S. Census region (Midwest, Northeast, South, and West). Additionally, we intended to enable estimation by urbanicity: city, suburban, town, and rural. Other targets of interest are districts with a high proportion of American Indian or Alaska Native (AIAN) students, and National Assessment of Educational Progress–Trial Urban District Assessment (NAEP-TUDA) districts.

Sampling frame. The sampling frame was constructed from the 2018–19 Common Core of Data (CCD) Local Education Agency Universe File, which contained the most recent public school district list available at the time of sampling. The 2018–19 CCD Universe File had 19,840 records with various characteristics about the districts. The final sampling frame consisted of 13,227 target school districts after the research team excluded nonregular districts; districts with no enrollment or no operational schools; districts labeled with closed, inactive, or future districts; and districts located outside of the 50 states and District of Columbia.

Stratification. To help reduce the sample variance for student-level estimates, the largest 140 districts are assigned to the certainty stratum. Additionally, districts with a high proportion (≥ 80%) of AIAN students and NAEP-TUDA districts were assigned to the certainty stratum. Overall, the certainty stratum included 275 districts. The remaining noncertainty districts were stratified by geographical areas (the 12 focal states or census regions) and urbanicity, which resulted in 64 strata.

Sample allocation. To reduce sample variance, we allocated the sample to strata proportionally to the square root of the total number of students within them based on a target of 1,500 responding districts (assuming a response rate of 70%). Additional cases were allocated among the strata with a respondent sample target of 60 districts for each focal state and 400 districts for each urbanicity category. See Table 1 for the allocated sample size for each of the 64 strata. Note that the stratum sample size shown in the table includes certainty districts.

Table 1. Allocated sample size by stratum

State Stratum	City	Suburban	Town	Rural	*Total*
California	61	88	25	24	*198*
Georgia	15	17	23	29	*84*
Illinois	10	60	20	25	*115*
Indiana	17	17	23	29	*86*
Kentucky	6	17	24	36	*83*
Massachusetts	11	40	19	36	*106*
Maryland	3	11	4	6	*24*
North Carolina	13	14	15	36	*78*
Oregon	14	17	21	30	*82*
Texas	38	44	35	64	*181*
Virginia	17	17	23	29	*86*
Washington	17	17	23	33	*90*
Rest of Midwest	47	124	105	195	*471*
Rest of Northeast	40	199	37	97	*373*
Rest of South	40	71	59	113	*283*
Rest of West	34	31	38	93	*196*
Total	*383*	*784*	*494*	*875*	*2,536*

Note. The stratum sample size shown in the table includes certainty districts.

Sample selection. To reduce the variance of student-based estimates, we sampled the noncertainty districts with probability proportional to size, with the measure of size being based on the number of students. In particular, the size measure was defined as $max\left(\sqrt{total\ enrollment}, 10\right)$. Setting a minimum measure of size limits the possibility of extreme weights, which will limit distortions in survey statistics.

The noncertainty sample was drawn systematically within each stratum. Districts in each stratum were sorted by a variable (percent of children in poverty) expected to correlate with survey variables of interest before sample selection. The sampling interval was computed as $SI^h = \sum^{i \in h} MOS^{hi} / n^h$, where n^h is the target sample size for stratum h, and MOS_{hi} is the measure of size for each district i in stratum h. We began with a starting value, $T = U[0,1] \cdot SI^h$, where $U[0, 1]$ was a random draw from a uniform distribution. For each stratum h, we incremented a running sum R based on MOS^{hi}, the measure of size for district i. If $R \geq T$, the district was selected into the sample, and T was incremented immediately by SI^h. We continued this process until n^h records were sampled from stratum h, and this processing was conducted independently for each stratum. The result of this sampling process was that all sample size targets were exactly met and each district's probability of selection was proportional to its size measure.

Survey Response Rate

Among the 2,536 sampled districts, 18 of them were determined to be out of scope during the data collection because the districts did not close schools to in-person schooling due to the COVID-19 epidemic or were no longer operational.

A completed survey was defined as having responses to a set of critical survey items. In total, there were 753 completed surveys and eight partially completed surveys.

The final survey response rate (29.9%) was calculated using response rate 1 from the American Association for Public Opinion Research (American Association for Public Opinion Research, 2016[1]): dividing the number of completes (*n* = 753 districts) by the number of all eligible sample cases (*n* = 2,518 districts).

Weighting Adjustments

The purpose of weighting was to enable unbiased estimates from the survey of public school districts. Without using weights in analyses, the estimates would be biased because the sample was not selected with equal selection probability and there were nonresponses. The final weights were produced by a multistep process, described next.

Base weights. The base weight for sampled districts was set to the inverse of the probability of selection. For districts in the certainty sample, their probability of selection was 1, so their base weight was also 1. For noncertainty districts i in stratum h, its probability of selection was set to $W_{hi} = n_h \cdot MOS_{hi} / \sum_{i \in h} MOS_{hi}$, and its base weight was the inverse of that.

[1] The American Association for Public Opinion Research. (2016). Standard definitions: Final dispositions of case codes and outcome rates for surveys. 9th edition. Washington, DC: Author. Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Nonresponse adjustments. The first step in our nonresponse adjustments was to identify variables that correlated with response propensities—variables that distinguished districts likely to respond from those not likely to respond. The variables we reviewed included census division or region, poverty status, focal state or region as used in the stratification, urbanicity, NAEP-TUDA participation, total students or teachers, percentage of AIAN students, certainty selection status, and grades covered. These variables were evaluated based on cross-tabulations with the response status and a logistic regression analysis. The cross-tabulations enabled us to see how many districts were in each cell and how the cells differed in terms of response rate. The logistic regression analysis confirmed which of the variables were important predictors of response status in a multivariate context. Based on this analysis, we selected three variables for forming nonresponse cells:

- Total enrollment (1–10,000; 10,001–50,000; 50,000+)
- Percentage of children in poverty (Deciles 1–6: 0–17.3% children in poverty; Deciles 7–10: 17.3%–100% children in poverty)
- Census division,[2] which divides a census region into smaller areas (Divisions 1 and 2 and Divisions 5, 6, and 7 were collapsed separately due to small cells)

For each of the nonresponse cells, we computed the weighted response rate excluding the 18 out-of-scope districts. The nonresponse adjusted weight was set to the base weight divided by cell response rate (i.e., the response rate for the cell in which each respondent fell). Note that this weight was computed only for respondents with a completed survey.

Weights trimming and calibration. Extreme variation in the survey weights can result in excessively large sampling variances when the variables of interest and the selection probabilities are not correlated, especially when the variation in the weights is a result of nonresponse adjustments. Therefore, we imposed a trimming strategy for excessively large weights. To reduce the variance of estimates generated with the weights, we trimmed weights falling outside a range of 3 times 0.95 of the interquartile range (IQR) of the median. After the weight trimming, the weights were recalibrated within the nonresponse cells. This trimming and recalibration process was repeated several times until all weights fell within the acceptable range (3 times the IQR of the median). Note that after trimming and recalibration, the sum of weights remained the same in total and by nonresponse cells.

The weights resulting from the trimming and recalibration process were the final weights and were intended for district-level analysis.

Evaluation of the final weights. As previously discussed, the purpose of weighting is to enable unbiased estimates using the final weights. Conducting analyses without applying the final weights may result in biased estimates. To evaluate the effectiveness of the final weights, we would need to compare the weighted estimates of variables of interest with population values; however, population values are not available for those variables. Therefore, we instead compared the weighted and unweighted estimates for variables available on the sampling frame with the population (frame) values. The comparison included

[2] Census division was only used as an additional categorization variable when the number of students in the district was 1–10,000. This was done so that nonresponse cell sizes were not made too small.

the three variables (total enrollment, percentage of children in poverty, and census division with collapsing due to small cells) used in the nonresponse weighting. As shown in Table 2, most of the unweighted estimates are different from the population values (reflecting the unequal sampling probabilities), while none of the weighted estimates are different from the population values. Note that the population values might include some potentially ineligible districts; however, the percentage of ineligible districts that remained in the sampling frame was expected to be very low, and it should not significantly affect the comparisons.

Table 2. Comparison of weighted and unweighted estimates of variables available on the sampling frame with the population (frame) values

Variable	Population Value[a]	Unweighted		Weighted	
		Estimate	Standard Error	Estimate	Standard Error
Mean enrollment	3,612.1	11,496.7*	596.5	3,521.5	130.9
Poverty (%)	16.3	17.7*	0.3	16.5	0.4
Census division (%)					
1 and 2	20.8	16.3*	0.2	21.7	0.9
3	20.8	19.9	0.8	20.9	1.2
4	15.3	8.5*	0.8	13.7	1.5
5, 6, and 7	23.5	32.5*	0.5	23.7	1.0
8	8.3	7.7	0.5	7.5	0.7
9	11.4	15.0*	0.6	12.4	0.9

Notes.

[a] Population value includes some potential ineligible districts. However, the percentage of ineligible districts remained in the sampling frame is expected to be low, and it should not significantly affect the comparisons.

* Statistically significant at the 5% level.

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

About the American Institutes for Research

Established in 1946, with headquarters in Washington, D.C., the American Institutes for Research (AIR) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance, both domestically and internationally, in the areas of education, health, and the workforce. **For more information, visit www.air.org.**

12930_10/20