AEP Data Note Technical Appendix

Christopher Joseph Doss, William R. Johnston

Sponsored by the Bill & Melinda Gates Foundation

For more information on this publication, visit www.rand.org/t/RR2575z1

Published (2018) by the RAND Corporation, Santa Monica, Calif.

Preface

This technical appendix provides additional information about the sample, data, and estimation strategy that was used for a series of American Educational Panels (AEP) Data Notes published by the RAND Corporation in 2018 and 2019. The Data Note series is intended to provide brief, incisive analyses of teacher and school leader survey results that may be of immediate interest to policymakers, practitioners, and researchers. The information in this technical appendix pertains to the first set of Data Notes, which will be released starting in 2018. If you are interested in using AEP data for your own analysis or reading other AEP-related publications, please email aep@rand.org.

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, financial literacy, and decisionmaking. This study was sponsored by The Bill and Melinda Gates Foundation, which is dedicated to improving the quality of life for individuals around the world. For more information and research on these and other related topics, please visit gatesfoundation.org.

More information about RAND can be found at www.rand.org. Questions about this appendix should be directed to cdoss@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

AEP Data Note Technical Appendix

The RAND American Educator Panels (AEP) consist of the American Teacher Panel (ATP) and American School Leader Panel (ASLP). These panels are nationally representative samples of K–12 public school educators. The ATP includes more than 20,000 teachers, and the ASLP includes more than 10,000 school leaders. Both groups respond to numerous online survey requests each year. The AEP began in 2014 and expanded significantly during the 2016–2017 school year. See Table 1 for a summary of the recruitment response rates during the 2016–2017 school year, as well as a summary of the response rates for the May 2017 survey.

Table 1. Panel Recruitment and Survey Response Rates

	ATP	ASLP
Recruitment		
Contacted for 2016–2017 panel recruitment effort	63,299	40,430
Newly enrolled in panel in 2016–2017	19,546	9,925
Enrollment rate	31%	25%
Survey administration		
Total panel size for May 2017 survey	20,986[a]	10,585
Complete responses	13,839[b]	4,515
Survey response rate	66%	43%

[a] The total panel size as of May 2017 is somewhat larger than the number of newly enrolled panel members because some participants joined the panels during earlier recruitment efforts.
[b] A teacher respondent had to complete at least 50 percent of the core survey to be considered a complete response for the ATP survey; for principals, the threshold was 33 percent of the ASLP survey.

Starting in 2013, AEP members have been recruited using probabilistic sampling methods. The AEP samples are designed to be of sufficient size to facilitate national analyses as well as analyses of prevalent subgroups at the national level (e.g., elementary school teachers, high school mathematics teachers, teachers in urban schools). Similarly, the panels are designed to permit analyses of the following geographic areas: Alabama, Arkansas, California, Florida, Georgia, Illinois, Kentucky, Louisiana, Maryland, Massachusetts, Mississippi, New Mexico, New York (New York State as a whole and New York City), North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia. One also can examine subgroups within these geographic areas (although there is lower precision for smaller groups). The AEP sample is not designed to permit analyses within geographic areas not listed above or among subgroups not specified above. For the initial AEP Data Note series, all analyses focus on national-level estimates.

Calibration Weights and Replicate Weights

A total of 81 weighting variables ensures that estimates reflect the national population of teachers and school leaders. One main weight was created to ensure that the sample is nationally representative, and a series of 80 replicate weights were calculated to produce the margin of error associated with an estimate. The main weight is calculated by first modeling response probabilities of teachers (or principals) across a wide variety of teacher (or principal) characteristics. The main weight is then calibrated so that the weighted sample matches the known national teacher or school leader population across these characteristics. Characteristics that factor into this process include descriptors at the individual level (e.g., gender, professional experience) and school level (e.g., school size, level, urbanicity, socioeconomic status). Replicate weights are calculated by removing 1/80th of the sample and repeating the process used to determine main weights on the remaining set of respondents. Replicate weights are used in calculating jackknife standard errors in our estimates.

The May 2017 Measurement Learning and Improvement Survey

In May 2017, RAND researchers administered the Measurement Learning and Improvement (MLI) Survey to the full ATP and ASLP samples. The MLI Survey was developed by the Bill & Melinda Gates Foundation (BMGF) in consultation with staff at RAND. Researchers provided feedback on question wording, format, and sequencing, with BMGF maintaining final editorial control on the survey items. The survey was designed to generate representative data on teacher and principal perspectives regarding educator preparation and working conditions, curriculum use and adherence to state standards, data use, and interventions to support students. Most items were developed by program staff at BMGF, with the exception of a set of publicly available questions from the Chicago Consortium on School Research's *5Essentials Survey*, which included items on leadership, parent-community ties, professional capacity, student-centered learning climate, and ambitious instruction (University of Chicago, 1994).

The data generated from the survey are intended to be used by analysts at RAND, BMGF, and numerous state education agencies (SEAs), with SEAs being able to compare the responses of educators from their states with a nationally representative comparison group. In addition to analyzing data from the core survey modules described above, SEAs were given the opportunity to develop customized modules for educators from their states. These modules averaged five minutes in length and were administered to educators in California, Colorado, Florida, Kentucky, Massachusetts, Mississippi, New York, North Carolina, and Oklahoma.

As shown in Table 1, the May 2017 MLI survey yielded 13,839 complete responses out of 20,986 invitations for teachers (66-percent response rate), and 4,515 out of 10,585 for school leaders (43-percent response rate). Table 2 provides weighted descriptive statistics for survey respondents.

Table 2. Weighted Descriptive Statistics

	ATP		ASLP	
	Mean	Standard Error	Mean	Standard Error
School characteristics				
Elementary school	0.489	0.004	0.591	0.017
Middle school	0.191	0.003	0.181	0.005
High school	0.269	0.002	0.181	0.001
Other types of schools	0.044	0.002	0.038	0.013
Total enrollment	848.670	9.580	566.690	8.560
Percentage Asian	5.115	0.297	4.444	0.674
Percentage Hispanic	23.703	0.615	22.516	0.825
Percentage black	15.610	0.477	14.913	0.931
Percentage white	50.933	0.740	53.316	0.230
Percentage Hawaiian/Pacific Islander	0.376	0.097	0.224	0.030
Percentage two or more races/ethnicities	3.192	0.095	3.389	0.209
Percentage English language learner	8.507	0.223	8.116	0.433
Percentage special education	13.073	0.077	13.632	0.580
Percentage free or reduced-price lunch	51.670	0.563	54.660	0.387
High poverty school (> 75% free or reduced-priced lunch)	0.224	0.009	0.262	0.002
Title I school	0.435	0.011	0.493	0.012
Urban school	0.289	0.001	0.263	0.003
Suburban school	0.385	0.003	0.321	0.001
Town school	0.117	0.000	0.129	0.001
Rural school	0.201	0.000	0.275	0.002
Educator characteristics				
Total years in role	14.438	0.134	7.111	0.111
Total years in role in school[a]	9.050	0.147	4.860	0.131
Male	0.232	0.006	0.489	0.010
Asian[a]	0.027	0.004	0.013	0.002
Hispanic[a]	0.074	0.007	0.081	0.010
Black[a]	0.071	0.007	0.120	0.015
White[a]	0.856	0.006	0.803	0.016
Hawaiian/Pacific Islander[a]	0.003	0.001	0.001	0.000
Other race/ethnicity[a]	0.030	0.004	0.012	0.002

NOTE: The ATP sample contains 13,839 observations. The ASLP sample contains 4,515 observations. School background characteristics were obtained from the Common Core of Data and are from the 2015–2016 school year. Means and standard errors were calculated using survey weights, which are calibrated to match the national averages for teachers and school leaders. The definition for *high poverty school* (greater than 75 percent free or reduced-price lunch) follows the definition set forth by the National Center for Education Statistics (National Center for Education Statistics, 2017). Educator characteristics are self-reported by the respondent. The rate of missingness in educator characteristics is about 2 percent and 5 percent in the teacher and principal samples, respectively.

[a] Variables were not used in the calculation of sampling weights.

Data Analysis

Primary Estimation Strategy

Our primary estimation strategy consists of simple weighted averages of the relevant survey responses among the full sample of respondents and/or subgroups of interest. To obtain simple national averages, we estimated the weighted means using the main weight described above. Replicate weights were used in calculating jackknife standard errors.[1] We also conducted analyses that contrasted survey responses across subgroups of interest.[2] In these analyses, to obtain point estimates and standard errors for each subgroup, the survey measure of interest was regressed on the indicator for the subgroup of interest as follows:

$$Y_{is} = \beta_0 + \beta_1 X_{is} + \varepsilon_{is} \quad (1)$$

Y_{is} represents the response of survey participant i in state s. X_{is} represents an indicator for whether the participant is a member of the subgroup of interest, X.[3] β_0 represents the average value of Y for respondents not in the subgroup of interest (the reference group). β_1 represents the differential response for members in the subgroup of interest. A test of the null hypothesis that $\beta_1 = 0$ will indicate whether the responses were statistically significantly different between the two groups. $\beta_0 + \beta_1$ provides the average response of participants in the subgroup of interest. ε_{is} is an individual stochastic error term. Survey weights and jackknife standard errors were used when reporting results from all regressions.

In some cases, the same questions were asked of teachers and principals, and the estimate of interest was the difference in responses between the two groups of educators. In these situations, the two data sets were combined, including each data set's overall weight and 80 replicate weights. Differences between groups were estimated using Equation 1, in which X_{is} represents an indicator for being a teacher and all 81 weights were included in the model. Some specification checks explored whether differences in response were affected by school characteristics, such as urbanicity, poverty level, grade level, or size. In these circumstances, the teacher indicator was interacted with the characteristic of interest. The magnitude and significance of the interaction

---

[1] Recent ATP and ASLP surveys are weighted so that a jackknife with 80 replication groups can be used. Each replication group is created by 1/80th of the sample (selected at random); the portion removed does not overlap across the groups. The weighting process that was used to create the main weights (wherein inverse probability weights were calibrated) is then applied to respondents within each replication group separately to create 80 sets of replication weights. A respondent who has been dropped from a specific replication group is given a weight of zero for weights corresponding to that group.

[2] Subgroup comparisons included comparing respondents from urban and non-urban schools and high-poverty and non–high-poverty schools.

[3] For illustration purposes, we use a comparison of two subgroups (e.g., teachers in urban schools versus teachers in non-urban schools). In cases involving more than two subgroups, the model would be adjusted to include additional X_{is} indicators of group membership, so that each group is being compared with an omitted referent.

term was used to judge whether the differences in responses were statistically significantly different among schools of a certain type.

Supplemental Analyses

To better understand whether the subgroup differences estimated in the primary model are confounded by characteristics of the schools or respondents, we also created a supplemental model that includes many additional covariates and a vector of state fixed effects. This supplemental model takes the following form (in which bold text indicates a vector of variables):

$$Y_{is} = \gamma_0 + \gamma_1 X_{is} + \boldsymbol{L_{is}}\gamma_2 + \boldsymbol{W_{is}}\gamma_3 + \alpha_s + \omega_{is} \quad (2)$$

In this model, we add a vector of school characteristics, $\boldsymbol{L_{is}}$, and a vector of individual characteristics, $\boldsymbol{W_{is}}$, and state fixed effects, α_s. School characteristics include indicators for school level (elementary school, middle school, high school, or other type of school); student enrollment; percentage enrollment by race or ethnicity (percentage Asian, percentage Hispanic, percentage black, percentage white, percentage Hawaiian/Pacific Islander, and percentage two or more ethnicities); percentage special education students; percentage free and reduced-price lunch students; an indicator for Title I eligibility; and an indicator of being an urban school. In models in which we explore response by high- and low-poverty schools, we replace the percentage free and reduced-price lunch enrollment with an indicator for being in a high-poverty school. We follow the National Center for Education Statistics' definition that a high-poverty school has 75 percent or more of students in a free or reduced-price lunch program (National Center for Education Statistics, 2017). In models in which we explore differences by urban and non-urban schools, we use only the indicator for urban schools. Respondent characteristics include the total number of years in the role (teacher or principal), the total number of years in the role at the current school, an indicator for male, and indicators for race or ethnicity (Asian, Hispanic, black, white, Hawaiian/Pacific Islander, other). State fixed effects are used to account for any state-level differences (i.e., factors that influence all educators in the same state) that are not observed in the available data.

We compare estimated β_1 from Equation 1 and estimated γ_1 from Equation 2 to understand the extent to which differences in responses by subgroup are confounded by school and individual characteristics. All reported relationships in the Data Notes were statistically significant in both specifications. In cases in which the two specifications yielded estimates that differ by more than 25 percent, we include both estimates in the appropriate Data Note. Any threshold used to report the two point estimates will be arbitrary. One challenge is that β_1 ranges from 10 percentage points to 40 percentage points. We believe the threshold of 25 percent represents differences in estimates that are potentially meaningful to the reader, regardless of the possible values of β_1.

Twenty states were oversampled in the design of the survey. When estimating effects on the principal sample, some replications of the jackknife standard errors could not be estimated

because of the small sample sizes in states that were not oversampled. In these instances, states that were not oversampled were aggregated into fixed effects by census region (Pacific, Mountain, West North Central, West South Central, East North Central, East South Central, New England, Middle Atlantic, and South Atlantic) to ensure the robustness of the standard errors. All models include survey weights and jackknife standard errors. To ensure robustness to clustered standard errors, regressions were also estimated in which the jackknife standard errors were replaced by standard errors clustered at the state level.

The MLI surveys were structured so that respondents were usually asked multiple questions within a domain (for example, distributed leadership in schools or social-emotional learning practices in schools); therefore, certain Data Notes discuss numerous comparisons of survey respondents and school or educator characteristics. Multiple comparisons within a similar domain might increase the risk of the type I error of "false discovery," which occurs when the null hypothesis of no association is rejected when, in fact, it is true. Therefore, we use the Benjamini-Hochberg multiple hypothesis testing correction, a widely used strategy in education research, to account for this increased risk of false discovery (U.S. Department of Education, 2017). All results reported to be statistically significant are robust to this correction.

Limitations

We acknowledge that all findings presented in the AEP Data Note series are based entirely on teacher and educator self-reports, which could be affected by social desirability bias and drive some responses to be more positive (e.g., teachers overreporting the extent to which they perform certain activities) or more negative (e.g., principals underreporting the resources or opportunities their districts provide as a way of explaining their own behavior). In addition, these biases can be correlated with the particular educator characteristic of interest in an analysis (i.e., the urbanicity of the school in which the educator works). Nonetheless, we believe these findings represent a reasonable demonstration of educators' perceptions of their working conditions and perspectives on their students, peers, and educational contexts. Results can help guide future research to uncover underlying causes for the patterns seen in the Data Notes, including social desirability bias.

Bibliography

National Center for Education Statistics, "Concentration of Public School Students Eligible for Free or Reduced-Price Lunch," webpage, March 2017. As of January 8, 2018: https://nces.ed.gov/programs/coe/indicator_clb.asp

University of Chicago, *5Essentials Survey*, Chicago, Ill.: University of Chicago Consortium on School Research, 1994.

U.S. Department of Education, *What Works Clearinghouse Procedures Handbook*, Version 4, Washington, D.C., 2017. As of May 29, 2018: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf