IES Institute of Education Sciences

Program for the International Assessment of Adult Competencies (PIAAC)

State-Level Estimation for Age and Education Groups Methodology Report

NCES 2022-050
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Statistics at IES

Program for the International Assessment of Adult Competencies (PIAAC)

State-Level Estimation for Age and Education Groups Methodology Report

June 2022

Jianzhu Li
Tom Krenzke
Weijia Ren
Leyla Mohadjer
Robert Fay
Andreea Erciulescu
Westat

Holly Xie
Stephen Provasnik
Project Officers
National Center for Education Statistics

NCES 2022-050
U.S. DEPARTMENT OF EDUCATION

IES Institute of Education Sciences

U.S. Department of Education
Miguel Cardona
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Statistics
Peggy Carr
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education
Potomac Center Plaza (PCP)
550 12th Street SW
Washington, DC 20202

June 2022

The NCES Home Page address is https://nces.ed.gov.
The NCES Publications and Products address is https://nces.ed.gov/pubsearch.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-12-C-0072 with Westat. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Li, J., Krenzke, T., Ren, W., Mohadjer, L., Fay, R., and Erciulescu, A. (2022). *Program for the International Assessment of Adult Competencies (PIAAC): State-Level Estimation for Age and Education Groups Methodology Report* (NCES 2022-050). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Publishing Office. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022050.

Content Contact
Holly Xie
(202) 245-8481
Holly.Xie@ed.gov

ACKNOWLEDGMENTS

The authors are grateful for the opportunity and guidance provided by the National Center for Education Statistics. At Westat, the efficient programming skills of Ying Chen, who led the statistical programming necessary for the production runs, was invaluable.

CONTENTS

CONTENTS—CONTINUED

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Program for the International Assessment of Adult Competencies (PIAAC) is a multicycle international survey of adult skills and competencies sponsored by the Organization for Economic Cooperation and Development (OECD). The survey examines a range of basic skills in the information age and assesses these adult skills consistently across participating countries. The first cycle of PIAAC included three rounds: 24 countries participated in 2011-12 (round 1), 9 additional countries participated in 2014-15 (round 2), and 5 additional countries participated in 2017-18 (round 3).

The United States participated in all three rounds of the first cycle of PIAAC. The round 1 (PIAAC 2012) survey design was consistent with the international requirements (OECD 2016). In round 2 (PIAAC 2014), a supplemental sample was drawn to enhance the round 1 sample (Hogan et al. 2016). The combined PIAAC 2012/2014 sample is nationally representative of the U.S. adult population 16-74 years old. The round 3 (PIAAC 2017) data collection had two core objectives. First, it was designed to produce a nationally representative sample of the U.S. adult population 16-74 years old. Second, the sample was designed to arrive at a large enough sample size that, when combined with the 2012 and 2014 samples, can produce small area estimates for counties in the United States.

This is the second methodology report on the PIAAC Small Area Estimation (SAE) published by the National Center for Education Statistics in PIAAC Cycle I. The first report was written to describe the methodology used for the purpose of creating model-based estimates for states and counties in the United States (Krenzke et al. 2020). The statistical modeling approach was used to produce model-based estimates, which are available to the public on the Skills Map website at https://nces.ed.gov/surveys/piaac/skillsmap/. In 2020, the statistical modeling approach produced four different state- and county-level estimates for adult literacy and numeracy proficiencies: an average score (on the PIAAC scale of 0 to 500) and the proportion of adults at or below Level 1, at Level 2, and at or above Level 3. In 2022, the statistical modeling approach was adapted to produce model-based estimates for six age groups and four education attainment groups for each state and the District of Columbia. This report is a follow-up to the initial report and describes the methodology for creating model-based estimates for age and education groups. While this report can be read without having to read the first report, it makes references to the content in the first report.

This report is organized into six sections. Section 1 provides an overview of the SAE process. Section 2 discusses the steps for preparing the survey estimates for input to the small area models. Section 3 provides details of the model estimation and prediction approach. Section 4 gives the technical description of the benchmarking approach. Section 5 presents various diagnostics and evaluation results, and the document concludes with a summary in section 6.

1. SMALL AREA ESTIMATION PROCESS OVERVIEW

Making effective evidence-based policies and laws relating to adult education requires sound research based on reliable data that are most relevant to the jurisdiction such as states and counties. As an international study involving more than 30 countries under the leadership of the Organization for Economic Cooperation and Development, the first cycle of the Program for the International Assessment of Adult Competencies (PIAAC) was designed to provide national estimates of the proficiency of adult literacy, numeracy, and problem-solving skills. In the United States, PIAAC is sponsored by the National Center for Education Statistics (NCES). The PIAAC survey provides high-quality national estimates through a multistage probability design with in-person data collections that include a screener questionnaire, a background questionnaire, and an assessment of adult skills. From 2012 to 2017, 12,330 U.S. adults ages 16-74 living in households were surveyed for PIAAC. Because the U.S. PIAAC sample size was too small to support the production of state and county estimates, small area estimation (SAE) methodology (Rao and Molina 2015) was used to produce model-based estimates of average scores for literacy and numeracy, and various proficiency levels. The model-based state and county estimates are available in the U.S. PIAAC Skills Map: State and County Indicators of Adult Literacy and Numeracy.[1] By using PIAAC survey data in conjunction with data from the American Community Survey (ACS), the Skills Map provides reliable estimates of adult literacy and numeracy skills in all 50 states, all 3,141 counties, and the District of Columbia. Of importance to its stakeholders, the Skills Map allows for the comparison of states and counties. The *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report* (Krenzke et al. 2020)[2] provides background on the PIAAC sample design and technical details about the model-based estimation process. Central to the small area estimation (SAE) process was an area-level bivariate Hierarchical Bayes (HB) linear threefold model for proportions, and a similar univariate model for averages.

The state and county estimates were produced so that policymakers can plan and allocate resources and target interventions as necessary at a more local level. PIAAC data are used by state adult education departments to plan interventions, allocate scarce resources, and provide information to the general public. To help further in targeting interventions, the set of model-based estimates in the Skills Map has been expanded to include state-level, model-based estimates for six age groups and four education groups. The age groups are 16-24, 25-34, 35-44, 45-54, 55-64, and 65-74. The education groups are less than high school, high school diploma or completion of the General Educational Development (GED) test, some college (no degree or attained associate's degree), and bachelor's degree

[1] Available at https://nces.ed.gov/surveys/piaac/skillsmap/.
[2] Available at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020225.

or higher. The state-level estimates of interest for the age and education groups (hereafter sometimes referred to as "groups") are the same as for the state and county estimates: proportions at or below Level 1, at Level 2, and at or above Level 3 and averages.

For the full 16-74 age range, there were 185 counties with sample data, including 15 small counties (counties with a sample size less than five, or just one segment), and there was an average of 67 respondents per county. The sample sizes available for survey estimates are critical input to area-level small area modeling. When split by age and education groups, the PIAAC sample sizes become smaller and have negative implications on the modeling process and results. Therefore, only state-level estimates have been generated using model-based estimation for the desired age groups and education groups; county-level estimation for age and education groups requires additional research. Table 1-1 shows the state-level sample size distributions for each age group and education attainment group. The mean state-level sample sizes range from 29 to 63 across the six age groups and from 46 to 81 for the four education attainment groups.

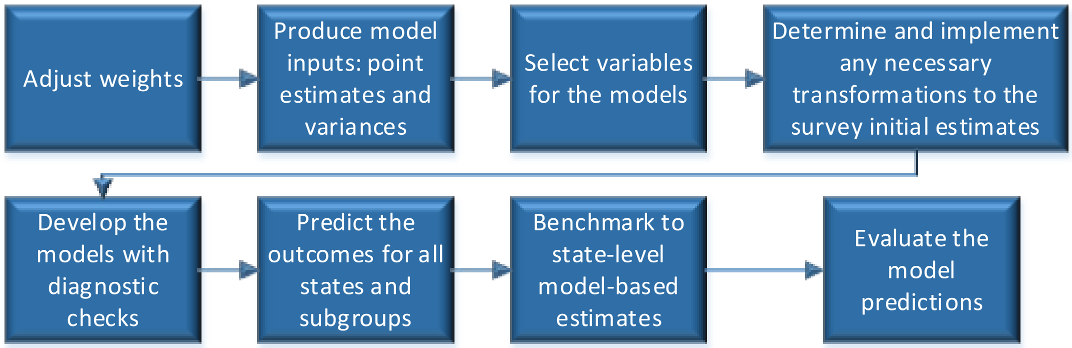
Table 1-1. State-level sample size distributions for age and education groups

Age and education groups	Number of states with sample	Number of respondents					
		Minimum	10th percentile	Median	90th percentile	Maximum	Mean
16-24	44	4	10	34	135	265	57
25-34	44	6	14	35	145	284	63
35-44	44	5	11	24	110	173	44
45-54	44	4	10	29	109	179	44
55-64	44	7	10	25	104	173	42
65-74	44	1	7	18	74	100	29
Less than high school	44	3	6	25	109	214	46
High school diploma or GED	44	5	13	53	202	288	81
Some college	44	13	17	55	180	351	79
Bachelor's degree or higher	44	11	17	39	183	329	75

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 1-1 provides a general overview of the process, which starts with a weight adjustment to calibrate the PIAAC sample weights to state-level demographic totals from the 2013-2017 ACS. State-level survey estimates on the key outcome quantities and associated variance estimates were produced using the adjusted weights.

Figure 1-1. Overview of the modeling process



Several regression models were used to select the key independent variables for the model, starting with the final covariates in the state and county models. Next, model fits were reviewed to determine if any necessary transformations were needed to the survey initial estimates to mitigate extreme proportions and unstable survey variances. Models were developed using the PIAAC initial state-level estimates for groups (after selected transformation for proportions) in conjunction with the selected covariates, and rigorous diagnostics checks were conducted. State-level model-based estimates were then derived for the groups. Model-based state-level estimates were constructed through weighted aggregation of the model-based state-level estimates for groups, and then compared against the model-based state-level estimates (from the state and county estimates process). The comparisons led to the development of a benchmarking process so that the aggregates based on the group estimates align with the state estimates. Many tables and graphs were prepared as part of the modeling process to assess the quality of the model-based estimates. Within the report, the results are illustrated for the proportion at or below Level 1 for literacy, for education attainment lower than a high school diploma. Results for other quantities of interest are available in the internal Supplemental Data Documentation and have been made available to NCES for internal review.

2. SURVEY ESTIMATES

The area-level modeling process for age and education groups began with computing state-level estimates. In section 2.1, a raking process is described to align the survey estimates to control totals from the ACS. Section 2.2 includes a summary of the methods used for the production of survey estimates and associated variances. Variance estimates are intended to account for the error associated with the item response theory modeling in addition to the sampling error. Ten plausible values (PVs) were generated for each respondent and a multiple imputation approach was used for calculating survey estimates and associated variance estimates. A variance smoothing process is presented with results in section 2.3. Refer to section 3 of the *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report* for details on accounting for informative nonresponse (literacy-related nonresponse) in the survey estimates.

2.1 Raking

The PIAAC final weight (SPFCWT0) was raked to the 2013-2017 ACS control totals, defined as the adult population size (age 16 to 74) for each raking dimension. The raking dimensions comprised age groups (16-24, 25-34, 35-44, 45-54, 55-64, 65-74); education levels (less than high school, high school or GED, some college, bachelor's degree or higher); gender (male, female); and race/ethnicity (Hispanic, non-Hispanic Black, non-Hispanic Others) within states.[3] The race/ethnicity groups were collapsed in some states, due to lack of sampled persons in these race/ethnicity groups. Table 2-1 shows the median of raking adjustment factors (ratio of raked weight to PIAAC final weight) and its interquartile range (IQR) by state. On average, the raking adjustment factors vary across states and may be much larger or smaller than 1 because the PIAAC sample was designed for national estimates, and primary sampling units (consisting of counties or groups of contiguous counties) were selected to represent areas inside and outside the state. To a lesser extent, the variation in raking factors is due to nonresponse adjustment and calibration that were conducted in the original weighting procedures. This raking process calibrated the weights of respondents to the state population estimates from the 2013-2017 ACS for all the sampled states consistently. In addition, 2013-2017 ACS state-level population totals and other state-level covariates were also available for the states without PIAAC sample data, and together

[3] The state-level control totals, as well as the state-level estimates that were used in modeling later, were extracted from the 2013-2017 ACS Summary File for all the U.S. counties. For the state and county estimates modeling process, the population was an approximation of the age 16-74 population from the 5-year ACS data. In this study for age and education groups, the population is the actual population estimates from the U.S. Census Bureau special tabulation of the ACS data.

with the ones for the states with PIAAC sample data, were used in the model-based estimates of the state by age or education groups; see section 3.

Table 2-1. Median and interquartile range (IQR) of raking adjustment factors, by state, sorted by descending median

State	Median	IQR
Ohio	3.77	3.35
Washington	3.25	2.40
Wisconsin	2.37	1.33
Maryland	1.98	0.72
Iowa	1.92	2.01
Massachusetts	1.79	1.56
Florida	1.55	0.84
South Carolina	1.47	1.12
Arizona	1.36	1.16
Michigan	1.28	0.59
Texas	1.21	0.22
New Jersey	1.15	0.50
Indiana	1.10	0.51
Alabama	1.08	1.03
New York	1.04	0.23
Mississippi	1.00	0.71
Louisiana	0.99	0.46
Oklahoma	0.98	0.36
California	0.94	0.08
Illinois	0.87	0.41
Tennessee	0.83	0.29
Utah	0.81	0.89
West Virginia	0.81	1.37
Connecticut	0.80	0.51
South Dakota	0.80	0.63
North Carolina	0.80	0.34
North Dakota	0.76	0.64
Arkansas	0.72	0.27
Georgia	0.71	0.22
Virginia	0.71	0.34
Colorado	0.66	0.17
Nevada	0.66	0.41
Pennsylvania	0.66	0.21
Minnesota	0.60	0.18
Kentucky	0.54	0.25
Nebraska	0.50	0.19

See note at end of table.

Table 2-1. Median and interquartile range (IQR) of raking adjustment factors, by state, sorted by descending median—Continued

State	Median	IQR
New Hampshire	0.48	0.17
Kansas	0.46	0.44
Montana	0.42	0.18
Hawaii	0.41	0.31
Delaware	0.41	0.25
Missouri	0.39	0.18
District of Columbia	0.25	0.19
Alaska	0.13	0.08

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

2.2 Generate Survey Estimates

The state-level survey estimates were generated for groups[4] defined in section 1. Similar to the approach used for the state and county estimates, the multiple imputation approach was implemented for calculating the survey estimates and the associated variance estimates, using the 10 PVs, which were discussed in the beginning of section 2. For the *m*-th plausible value for group *g,* in state *j*, the survey estimate was computed as

$$\hat{y}_{jg}^{(m)} = \sum_{a=1}^{n_{jg}} w_{jga} y_{jga}^{(m)} / \sum_{a=1}^{n_{jg}} w_{jga}, \tag{2a}$$

where w_{jga} is the raked weight for person *a* in state *j*, group *g*; $y_{jga}^{(m)}$ is the proficiency score (for average) or an indicator variable for the proficiency level (for proportions); and n_{jg} is the number of cases in state *j*, group *g*.

Then the state-level survey estimate for groups ($\hat{y}_{jg}$) was calculated as

$$\hat{y}_{jg} = \frac{1}{10} \sum_{m=1}^{10} \hat{y}_{jg}^{(m)}. \tag{2b}$$

The Multiple Imputation estimate of the variance is

$$\hat{\sigma}_{jg}^2 = \hat{\sigma}_{Wjg}^2 + \left(\frac{11}{10}\right) \hat{\sigma}_{Bjg}^2, \tag{2c}$$

[4] There are 264 state by age groups (calculated as 44 states multiplied by six age groups) and 176 state by education groups (calculated as 44 states multiplied by four education groups) in total.

where $\hat{\sigma}^2_{Wjg}$ is the within-imputation variance and $\hat{\sigma}^2_{Bjg}$ is the between-imputation variance. The within-imputation variance component was computed as the average of the sampling variance for each of the 10 PVs

$$\hat{\sigma}^2_{Wjg} = \left(\sum_{m=1}^{10} \hat{v}_{jg}^{(m)}\right)/10, \tag{2d}$$

where $\hat{v}_{jg}^{(m)}$ is the sampling variance of the estimated mean or proportion for plausible value *m*. The between-imputation component was calculated as

$$\hat{\sigma}^2_{Bjg} = \left[\sum_{m=1}^{10} \left(y_{jg}^{(m)} - \hat{y}_{jg}\right)^2\right]/9 \, . \tag{2e}$$

Sampling variances were calculated using the Taylor series method (Wolter 2007), with secondary sampling units (SSUs) as variance units (clusters) due to the limited number of primary sampling units (PSUs). SSUs were generally groups of U.S. Census Bureau (Census) blocks. While the between PSU variance is not explicitly accounted for in this approach, a portion of the between PSU variance is included in states with more than one sampled PSU. This solution achieves greater stability of the variance estimate at the cost of some downward bias. But the alternative of estimating variance using PSUs as clusters would have yielded a less stable variance estimate because of the small number of PSUs in each state. In addition, no variance estimate would result in states with only one sample PSU.

The direct variance estimate could not be computed for one state and age group that only had one SSU with PIAAC data. The models that are described in section 2.3 were used to predict the variances for this group. The remaining 263 age groups and all the 176 education groups had at least two SSUs and sample sizes of four or more. Table 2-2 shows the distribution of the proportion of variance for survey estimates attributed to imputation error, as measured through multiple imputation (i.e., between-imputation variance) across the 263 age groups and 176 education groups with at least two SSUs, for both literacy and numeracy components. For literacy skills age groups, multiple imputation contributes on average 14 percent of total variance for the average score, 22 percent of total variance for the proportion at or below Level 1, 29 percent of total variance for the proportion at Level 2, and 16 percent of total variance for the proportion at Level 3 and above. Across age groups, the contribution to the total variance from multiple imputation ranges from nearly 0 percent to 80 percent. For literacy skills education groups, multiple imputation contributes, on average, 16 percent of total variance for the average score, 25 percent of total variance for the proportion at or below Level 1, 28 percent of total variance for the proportion at Level 2, and 18 percent of total variance for the proportion at Level 3 and above. Across education groups, the contribution to the total variance from multiple imputation ranges from nearly 1 percent to 70 percent. The distribution of the proportion for numeracy is similar to that for literacy.

Table 2-2. Distribution of the proportion of variance associated with multiple imputation for survey estimates across groups, across states: 2012/2014/2017 PIAAC

Proficiency component	Group	Proportion of variance due to multiple imputation for	N[1]	Mean	Minimum	Maximum	Std dev
Literacy	Age	Average score	263	0.14	0.00	0.80	0.111
Literacy	Age	Proportion at or below Level 1	261	0.22	0.01	0.62	0.136
Literacy	Age	Proportion at Level 2	263	0.29	0.04	0.61	0.112
Literacy	Age	Proportion at Level 3 and above	263	0.16	0.01	0.59	0.105
Literacy	Education	Average score	176	0.16	0.01	0.68	0.121
Literacy	Education	Proportion at or below Level 1	175	0.25	0.02	0.69	0.140
Literacy	Education	Proportion at Level 2	175	0.28	0.05	0.61	0.125
Literacy	Education	Proportion at Level 3 and above	174	0.18	0.01	0.62	0.125
Numeracy	Age	Average score	263	0.16	0.02	0.83	0.115
Numeracy	Age	Proportion at or below Level 1	263	0.22	0.02	0.66	0.134
Numeracy	Age	Proportion at Level 2	263	0.30	0.05	0.72	0.121
Numeracy	Age	Proportion at Level 3 and above	263	0.17	0.01	0.63	0.108
Numeracy	Education	Average score	176	0.19	0.02	0.67	0.114
Numeracy	Education	Proportion at or below Level 1	176	0.23	0.03	0.77	0.127
Numeracy	Education	Proportion at Level 2	176	0.31	0.05	0.67	0.127
Numeracy	Education	Proportion at Level 3 and above	173	0.18	0.01	0.63	0.118

[1]There is one age group having only one secondary sampling unit (SSU); thus, the variance cannot be estimated. For specific proficiency component, some age or education groups have no respondents with scores at certain levels, thus variance is 0 for that group and proportion of variance associated with multiple imputation is missing.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

2.3 Variance Smoothing for Averages

Similar to the state and county model-based estimates, variance smoothing was implemented to predict the unknown true variance. Due to the transformation applied to the proportion survey estimates (described in section 3.2.1), variance smoothing was only applicable to the average proficiency scores. The design effect associated with weight variation and clustering within the small areas were considered in the variance smoothing. Specifically, the variance is smoothed by fitting a weighted least squares model as shown below:

$$ln(\hat{\sigma}_{jg}^2) = \beta_0 + \beta_1 ln(C_{jg}) + \beta_2 ln(B_{jg}) + \beta_3 ln(\hat{\sigma}_{y_{jg}}^2) + \epsilon,$$

where $ln(\hat{\sigma}_{jg}^2)$ is the natural log of the multiple imputation variance for each group g in state j, $ln(C_{jg})$ is the natural log of the number of SSUs in each group, $ln(B_{jg})$ is the natural log of the average cluster size for each group g in state j, and $ln(\hat{\sigma}_{y_{jg}}^2)$ is the natural log of estimated population variance of the literacy/numeracy scores among each group g in state j. The model is weighted by $C_{jg} - 1$. The exponentiation of the predicted value from this model would be the smoothed variance. Two models were fitted, one for the age groups and the other for the education group. It should be noted that groups with fewer than four clusters were excluded from the smoothing process, so the age smoothing model is estimated based on 258 age groups and the education smoothing model is based on 173 education groups. After the models were fitted, the smoothed variances were derived for all 264 age groups and 176 education groups. In later sections, the smoothed variances are denoted by σ_{jg}^2. Table 2-3 provides the parameter estimates for the smoothing process for literacy and numeracy average scores. The smoothing approach helped reduce the standard deviation of the variances in most proficiency components and age and education groups.

Table 2-3. Summary of average proficiency score variance estimates prior to and after smoothing: 2012/2014/2017 PIAAC

Proficiency component	Group	Stage	Number of states[1]	Minimum	Median	Mean	Maximum	Standard deviation
Literacy	Age 16-24	Survey estimate	44	15.89	86.92	184.40	1021.04	219.02
		Smoothed	44	27.27	98.13	142.75	726.60	145.08
	Age 25-34	Survey estimate	44	28.92	93.38	143.75	728.51	141.92
		Smoothed	44	37.68	95.44	142.03	538.77	110.62
	Age 35-44	Survey estimate	44	40.89	140.56	204.69	1113.33	233.49
		Smoothed	44	45.83	139.69	181.06	747.41	153.42
	Age 45-54	Survey estimate	44	40.17	159.30	261.47	1996.90	332.96
		Smoothed	44	45.73	176.74	233.67	1417.63	279.27
	Age 55-64	Survey estimate	44	23.00	112.15	183.94	1283.05	210.93
		Smoothed	44	38.17	130.61	181.66	612.66	145.52
	Age 65-74	Survey estimate	43	46.45	172.30	205.92	807.72	157.60
		Smoothed	43	59.29	180.98	191.40	652.23	117.86
	Less than high school	Survey estimate	44	31.61	137.74	290.45	1802.32	351.25
		Smoothed	44	32.37	124.04	257.16	1587.14	326.06
	High school	Survey estimate	44	13.85	56.61	110.31	577.35	119.54
		Smoothed	44	18.19	70.33	103.81	551.59	109.64
	Some college	Survey estimate	44	8.63	48.09	87.89	451.72	86.13
		Smoothed	44	18.09	58.12	86.99	381.38	83.67
	Bachelor's degree or higher	Survey estimate	44	13.37	49.06	67.27	219.42	54.03
		Smoothed	44	15.48	59.38	72.36	267.80	60.43

See notes at end of table.

Table 2-3. Summary of average proficiency score variance estimates prior to and after smoothing: 2012/2014/2017 PIAAC—Continued

Proficiency component	Group	Stage	Number of states[1]	Minimum	Median	Mean	Maximum	Standard deviation
Numeracy	Age 16-24	Survey estimate	44	27.81	113.40	178.32	856.20	179.76
		Smoothed	44	27.45	97.28	156.73	766.15	156.48
	Age 25-34	Survey estimate	44	24.14	106.53	149.49	851.25	149.76
		Smoothed	44	29.16	118.66	152.80	525.66	119.76
	Age 35-44	Survey estimate	44	39.28	170.32	198.82	955.14	185.56
		Smoothed	44	42.80	143.10	188.29	675.08	148.44
	Age 45-54	Survey estimate	44	43.80	159.49	295.50	2645.48	425.26
		Smoothed	44	48.72	152.25	256.94	1620.78	324.17
	Age 55-64	Survey estimate	44	29.60	130.82	183.96	730.94	164.28
		Smoothed	44	35.46	158.01	192.02	657.57	148.26
	Age 65-74	Survey estimate	43	43.23	209.37	222.98	739.16	155.09
		Smoothed	43	61.41	151.07	207.42	761.21	148.10
	Less than high school	Survey estimate	44	45.24	157.21	295.72	1272.95	299.07
		Smoothed	44	27.52	158.47	261.68	1210.95	273.67
	High school	Survey estimate	44	19.55	76.83	125.15	770.47	139.54
		Smoothed	44	19.74	74.91	113.28	710.47	122.61
	Some college	Survey estimate	44	15.90	58.77	99.25	460.89	105.89
		Smoothed	44	17.78	70.58	93.37	379.69	80.68
	Bachelor's degree or higher	Survey estimate	44	9.40	65.09	85.86	375.84	78.38
		Smoothed	44	13.73	77.37	95.37	384.26	79.67

[1] There is one age group having only one secondary sampling unit (SSU), and it is excluded from this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

3. MODEL ESTIMATION AND PREDICTION

The next step in the production of the model-based group estimates fitted area-level small area models to the state-level survey estimates described in the previous section. Separate models were developed for each group. In general, the models were simplified relative to the models for the state and county estimates for the U.S. PIAAC Skills Map. With one exception, only the seven variables[5] that had been selected for the state and county modeling were considered for the age and education group models. Section 3.1 describes the modifications to these variables as candidates for the group modeling. The set of candidate variables was then further restricted to those contributing the most to the overall fit. A final set of variables for the age-group modeling was determined from informal initial regression fits followed by an assessment of nearby alternative sets of variables using the Akaike information criterion (Akaike 1974). A final set for the education group modeling was determined by the same process. In section 3.2, the model specifications are described for proportions, including use of the arcsine transformation, and for averages. Details of the model estimation and production of predicted values and measures of precision then follow.

3.1 Covariate Formation

The production of state and county estimates included selecting model covariates out of a large set of possible covariates. To produce the new state-level estimates for age and education groups, the general strategy was to build on the previous research to the extent feasible, but also to modify the details where necessary. Accordingly, consideration of covariates for the age- and education-group estimates began with the seven covariates that had been chosen for the state and county estimates. For the educational groups, additional covariates were also considered, namely covariates giving the percentage of population in the six age groups within the educational group.

3.1.1 Variable Modification

To produce estimates for the six age groups, 16-24, 25-34, 35-44, 45-54, 55-64, and 65-74, the covariates used for the state and county estimates were modified to reflect age-specific rates, rather than the original overall rate. For example, for the 16-24 age-group estimation, the proportion with less than a high school education was replaced by the proportion of individuals age 16-24 with less than a high

[5] The seven covariates are related to educational attainment, poverty, race/ethnicity, health insurance coverage, and occupation (service industry).

school education. In general, age-specific versions were used for each of the covariates that had been used to produce the state and county estimates.

A similar strategy was used for the four education groups: less than high school, high school diploma or GED, some college (no degree or attained associate's degree), and bachelor's degree or higher. In this case, education-specific rates were produced. However, the educational attainment covariates, which were important predictors in the state and county estimates, could no longer be used to produce estimates by education group. For example, the proportion with less than high school was no longer suitable for predicting estimates for the education group with less than high school, because the proportion would be 1.00 in each state. Instead, variables summarizing the age distribution within each education group were introduced. For example, one such covariate for use with the less than high school diploma education group was the proportion in ages 16-24 of the total in the less than high school education population. Similarly, covariates for the proportions in ages 25-34, in ages 35-44, in ages 45-54, and in ages 55-64 were defined for each education group.

As noted in the introduction, the reduction in sample sizes for each age or education group placed additional methodological constraints on the small area models. The survey estimates serving as input into the models were constructed at the state level rather than at the county level, to achieve some degree of stability. However, this change substantially reduced the number of geographic areas to build the models. To avoid fitting models with a high number of parameters to a limited number of areas, further simplifications to the models were required. One simplification was to reduce the set of covariates with the goal of cutting the number of covariates to between three and five, or roughly 10 percent of the number of states with PIAAC observations. A second simplification was to fit only univariate models for the proportions at Level 1 or below, or for the proportions at Level 3 or above, rather than a multivariate model for these variables, because preliminary modeling suggested that the multivariate models were relatively unstable. Finally, the Census division-level random effects were eliminated. Preliminary modeling consistently showed very low yields for these random effects for the age- or education-group model estimates. More details about the final model specifications are presented in section 3.2.

3.1.2 Final Reduction of Covariates

A list of the candidate covariates is provided in table 3-1, which indicates in the initial group column the group(s) for which the models potentially could include each covariate. As noted in the previous subsection, the education variables were only considered for the age group models, whereas the age covariates at the bottom of the table were only considered for the education groups. The final group column indicates the final determinations to be described below.

Table 3-1. List of covariates for the small area models, initially considered for the state model and finally selected

Covariate	Label	Initial group	Final group
Percentage of population with less than high school education	LH	Age	Age
Percentage of population with more than high school education	MH	Age	Age
Percentage of population below 100 percent of the poverty line	Poverty	Age, Education	Age, Education
Percentage of Black or African American population	Black	Age, Education	
Percentage of Hispanic population	Hispanic	Age, Education	
Percentage of civilian noninstitutionalized population who has no health insurance coverage	Insurance	Age, Education	Education
Percentage of population age 16 and over with service occupations	Service	Age, Education	Age
Percentage of population age 16-24	Age_16_24	Education	Education
Percentage of population age 25-34	Age_25_34	Education	
Percentage of population age 35-44	Age_35_44	Education	
Percentage of population age 45-54	Age_45_54	Education	
Percentage of population age 55-64	Age_55_64	Education	
Percentage of population age 65-74	Age_65_74	Education	

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

To search for a possible further reduction in number of covariates for the age groups and education groups, fixed-effect regression models weighted by the state sample size were fitted to the six age-specific models for the proportion at or below Level 1 literacy (P1 Lit), the six age-specific models for the proportion at or above Level 3 literacy (P3 Lit), and six age-specific models for mean literacy score (Lit mean)—18 models in all—as well as the 18 analogous models for numeracy. A corresponding set of 24 fixed-effect regression models was fitted to the education-specific rates, including the new covariates for proportions in each age category for specific education groups. The frequency of statistically significant results at the .10 level was used as a guide to which covariates were making the strongest contribution. For age-specific rates, the covariates that appeared to have the most effect were (1) percentage in service occupations, (2) less than high school, (3) more than high school, and (4) poverty. For education groups, the results appeared to favor the following covariates: (1) no health insurance, (2) poverty, (3) percentage of population age 35-44, and (4) percentage of population age 55-64.

A second step, still using fixed-effect regressions weighted by the state sample size, compared these initial suggestions choosing similar covariates with a more formal approach, namely the

Akaike information criterion (AIC) (Akaike 1974). The AIC allows the comparison of different models, trading off the explanation offered by the selection of covariates against the cost of estimating the coefficients of the model. AIC takes the form

$$AIC = 2k - 2\ln(\hat{L}),$$

where k is the number of parameters (in this case, the number of covariates plus 1 for the intercept) and $\hat{L}$ is the estimated maximum of the likelihood function. The models with the lowest AIC are generally preferred, though differences of 2 or even 4 can be attributed to chance. Although not entirely justified, the assumption of statistical independence permits adding the six values of AIC across age groups as an assessment of the overall success of models based on a set of covariates.

Table 3-2 presents sums of AIC scores for several age-group models, ranked in increasing order of AIC. Overall, the AIC prefers the same model as suggested by the initial regression models. The results indicate that the covariate more than high school contributes little to modeling the proportion at or below Level 1 literacy, but including this covariate is especially important in modeling the proportion at or above Level 3 numeracy. Consistent with the AIC results, the combination of variables in the first row of the table was selected to model the age-group estimates.

Table 3-2. Total of Akaike information criterion (AIC) measures over six age groups for fitting different outcomes and the total across all six outcomes

Model	P1 Lit	P1 Num	P3 Lit	P3 Num	Lit mean	Num mean	Total
Service + LH + MH + Poverty	-384.8	-371.5	-378.7	-350.8	2111.8	2141.0	2767.0
LH + MH + Poverty	-383.6	-372.0	-378.9	-349.7	2114.7	2145.8	2776.2
Service + LH + Poverty	-392.3	-368.6	-378.1	-329.2	2102.9	2150.5	2785.1
LH + Poverty	-389.4	-369.1	-377.9	-327.7	2106.7	2155.3	2797.9
Service + LH + MH + Insurance + Poverty	-376.1	-365.3	-372.8	-345.6	2119.9	2142.5	2802.4
Service + LH + Insurance + Poverty	-382.8	-364.0	-377.3	-329.8	2109.5	2147.1	2802.7
LH + MH + Poverty + Insurance	-374.8	-366.9	-375.4	-345.6	2121.9	2146.0	2805.2
Service + LH + MH	-382.3	-349.4	-373.5	-340.0	2112.3	2153.8	2820.8
Service + LH + MH + Insurance	-374.2	-344.0	-369.6	-335.2	2119.1	2154.1	2850.2
Service + LH	-391.9	-335.6	-366.6	-291.5	2102.1	2178.8	2895.3
Service + LH + Insurance	-382.2	-333.5	-368.2	-294.6	2108.1	2171.8	2901.2

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Similarly, table 3-3 presents sums of AIC scores for several education-group models, ranked in increasing order of AIC. The AIC prefers a model replacing the proportions for ages 35-44 and 55-64 by the proportion for ages 16-24. The AIC for this model is either the best or within range of the best AIC for each of the literacy and numeracy measures. Here, too, the variables in the first row were selected for the education-group estimates.

Table 3-3. Total of Akaike information criterion (AIC) measures over four education groups for fitting different outcomes and the total across all six outcomes

Model	P1 Lit	P1 Num	P3 Lit	P3 Num	Lit mean	Num mean	Total
Insurance + Poverty + Age_16_24	-298.3	-306.3	-313.7	-275.2	1356.8	1357.6	1520.9
Insurance + Poverty + Age _55_64 + Age _35_44	-292.3	-302.8	-308.7	-277.8	1359.0	1356.7	1534.2
Insurance + Poverty + Age _55_64	-286.0	-306.6	-312.9	-277.6	1364.0	1355.5	1536.5
Insurance + Poverty + Age _35_44	-282.5	-291.5	-299.7	-265.7	1373.5	1379.6	1613.8
Poverty + Age _16_24	-292.4	-288.1	-286.4	-255.6	1375.6	1387.5	1640.5
Poverty + Age _16_24 + Black	-286.0	-287.6	-283.7	-259.6	1381.2	1386.8	1651.0
Service + Poverty + Age _16_24	-291.1	-285.6	-280.2	-252.0	1379.3	1393.6	1663.9
Service + Poverty + Age _16_24 + Black	-283.9	-284.2	-277.9	-256.1	1385.2	1392.7	1675.8
Age _16_24 + Black	-273.8	-278.9	-277.0	-249.2	1392.2	1393.2	1706.4
Service + Age _16_24 + Black	-273.2	-276.2	-273.2	-249.5	1393.0	1397.7	1718.6
Poverty + Black	-256.3	-286.1	-278.1	-254.8	1409.3	1409.5	1743.5
Service + Poverty + Black	-252.3	-282.5	-272.8	-252.0	1414.7	1415.8	1770.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

3.2 Modeling

The SAE model specifications for proportions and averages for literacy and numeracy are described in sections 3.2.1 and 3.2.2, respectively.

3.2.1 Modeling Proportions – Area-Level Univariate HB Linear Model

For each age or education group of interest, an area-level HB linear model was used for estimating PIAAC proportions, with states being the areas. The two proportions represented by Level 1 and below (P1) and Level 3 and above (P3) were modeled for each group independently. The results were then combined to estimate the proportion in Level 2 (P2).

3.2.1.1 Arcsine Square Root Transformation

The arcsine square root transformed survey estimates were used in the HB models as the dependent variables. For a state *j*, group (age or education) *g*, the transformed survey estimate is defined as

$$y_{jg} := sin^{-1}\sqrt{\hat{p}_{jg}},$$

where $\hat{p}_{jg}$ is the survey estimate for proportion. The arcsine square root transformation was applied to the proportions for the purpose of stabilizing their variances. The estimated variance of y_{jg} is

$$\frac{\hat{V}_{jg}}{4\hat{p}_{jg}(1-\hat{p}_{jg})},$$

where $\hat{V}_{jg}$ is the estimated variance of $\hat{p}_{jg}$. Note that this fraction can be estimated by $\frac{1}{4n_{e,jg}}$, where $n_{e,jg}$ is the area-level PIAAC effective sample size. However, this quantity is not directly defined for areas with estimated proportions $\hat{p}_{jg}$ being 0 or 1, or for areas with undefined variance estimates $\hat{V}_{jg}$. To overcome this challenge, the area-level effective sample sizes can be approximated using Kish's formula

$$\tilde{n}_{e,jg} \approx \frac{\left(\sum_{a\in j\times g} w_{ajg}\right)^2}{\sum_{a\in j\times g}\left(w_{ajg}\right)^2},$$

where w_{ajg} is the raked weight associated with the individual *a* in state *j*, group *g*. Hence, the variance estimate associated with y_{jg} is defined as

$$\sigma_{jg}^2 := \frac{1}{4\tilde{n}_{e,jg}} \approx \frac{1}{4}\frac{\sum_{a\in j\times g}\left(w_{ajg}\right)^2}{\left(\sum_{a\in j\times g} w_{ajg}\right)^2}.$$

3.2.1.2 Model Specification

The HB model includes a sampling model and a linking model, for each group *g* at the state level, as follows:

$$\begin{aligned} y_{jg} &\sim N(\theta_{jg}, \sigma_{jg}^2) \\ \theta_{jg} &\sim x_{jg}'\beta_g + v_{jg}, \end{aligned}$$

where y_{jg} are the arcsine square root transformed survey estimates for proportions at or below Level 1 or at or above Level 3, x'_{jg} is a vector of covariates, β_g is a vector of coefficients, and v_{jg} is a state-level random effect. The estimated variance σ^2_{jg} is the approximated variance described above and it is treated as fixed and known. Independent priors are assumed for the regression coefficients and the random effects. Specifically, it is assumed $\beta_g \sim N(0,10000)$ component-wise, where the normal distribution specification uses the mean and the variance. It is also assumed that the random effect v_{jg} follows a normal distribution with mean zero and variance $\sigma^2_{v,g}$. A Cauchy distribution with hyperparameters 0 and 5 is adopted as prior distribution for the standard deviation parameter $\sigma_{v,g}$. These choices of prior distributions ensure that they provide little information to the estimation of model parameters, whose posterior distributions mainly depend on the data.

3.2.2 Modeling Averages – Area-Level Univariate HB Linear Model

Similar to the HB models for proportions, for each age or education group of interest, an area-level HB linear model was used for estimating PIAAC averages, with states being the areas. The model for averages is specified as follows:

$$\begin{aligned} y_{jg} &\sim N(\theta_{jg}, \sigma^2_{jg}) \\ \theta_{jg} &\sim x'_{jg}\beta_g + v_{jg}, \end{aligned}$$

where y_{jg} is the survey estimate of average literacy or numeracy scores in group g at the state level, with associated estimated variance σ^2_{jg} , x'_{jg} is a vector of covariates, β_g is a vector of coefficients, and v_{jg} is the state-level random effect. The estimated variances σ^2_{jg} are the result of smoothing functions of the variances for the survey estimates (see details in section 2.3) and treated as fixed and known in the HB model. It is also assumed that the random effect v_{jg} follows a normal distribution,

$$v_{jg} \sim N\left(0, \sigma^2_{v,g}\right).$$

Independent priors are assumed for the regression coefficients and the random effects variance. Specifically, it is assumed $\beta_g \sim N(0,1000000)$ component-wise, where the normal distribution specification uses the mean and the variance. The variance of the random effects, $\sigma^2_{v,g}$, is assumed to follow a uniform distribution over a wide range, 0 to 10,000.

3.2.3 Model Estimation

The univariate HB models were fitted for estimating proportions and averages at the state level for each group of interest and for literacy and numeracy, respectively, based on the data from 44 states in the PIAAC sample. Model fitting was carried out using a Markov chain Monte Carlo (MCMC) method. The RStan software was employed for this purpose. Three independent Markov chains were processed to facilitate the calculation of Monte Carlo standard errors (see Gelman and Rubin 1992; Rao 2003). The procedure started with three sets of initial values corresponding to the three independent chains, and then updated all the parameter values in each chain. After tossing 20,000 iterations in the warm-up period, 15,000 further iterations were produced for each of the three chains and were "thinned" by taking 1 in every 10. Thus, over the three chains, a total of 4,500 iterations remained. These 4,500 final iterations (referred to as MCMC samples) then simulated the posterior distributions of all the model parameters. The parameters β and the variance of the random effect are shown in table 3-4 for the HB model for the less than high school group and for literacy proportion at or below Level 1.

Table 3-4. Regression coefficients and components of the variance-covariance matrices of random effects for the final Hierarchical Bayes (HB) model: for the less than high school group and for literacy proportion at or below Level 1: 2012/2014/2017 PIAAC

Parameter	HB mean	HB standard deviation	Median	95 percent credible interval: Lower bound	95 percent credible interval: Upper bound
Intercept	0.39	0.30	0.39	-0.19	0.97
Insurance	0.36	0.29	0.37	-0.24	0.91
Poverty	1.90	0.64	1.89	0.66	3.16
Age_16_24	-0.55	0.41	-0.56	-1.34	0.27
$\sigma_{v,g}$	0.05	0.03	0.05	0.00	0.12

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

3.2.4 Predicted Values

Once the final model was processed and the model parameters estimated, the next step was to estimate the proportions and averages for the sampled states and the nonsampled states.

The posterior mean $\hat{\theta}_{jg}^{HB}$ of a sampled state for proportions or averages is produced as follows:

$$\hat{\theta}_{jg}^{HB} = \frac{\sum_{b=1}^{4,500} \theta_{jg}^{(b)}}{4,500}, \tag{3a}$$

where, for averages, the value of $\theta_{jg}^{(b)}$ for MCMC sample b is obtained from

$$\theta_{jg}^{(b)} = x'_{jg}\beta_g{}^{(b)} + v_{jg}^{(b)}, \tag{3b}$$

and for proportions, the value of $\theta_{jg}^{(b)}$ for MCMC sample b is obtained from converting the estimates from the MCMC samples back to the form of proportions

$$\theta_{jg}^{(b)} = \left(\sin\left(x'_{jg}\beta_g{}^{(b)} + v_{jg}^{(b)}\right)\right)^2. \tag{3b}$$

For all the nonsampled states, the values of $v_{jg}^{(b)}$ were not available. Following Rao (2003), $v_{jg}^{(b)}$ was drawn from $N\left(0, \sigma_{v,g}^{2\ (b)}\right)$.

3.2.5 Measures of Precision for the Model-Based Estimates

The primary measure of precision reported for each state and group model-based estimate is its credible interval. The 95 percent credible intervals for the state estimates $\hat{\theta}_{jg}^{HB}$ were computed by calculating the 2.5 percent (lower bound) and 97.5 percent (upper bound) quantiles of the 4,500 MCMC samples $\theta_{jg}^{(b)}$, b = 1,…,4,500 that simulated the posterior distributions. Since these posterior distributions are skewed, the credible intervals are nonsymmetric around the estimate.

An alternative measure of uncertainty is the coefficient of variation (CV). The CV of the HB estimate for state *j* is computed as

$$CV_{jg} = \frac{\sqrt{Var\left(\hat{\theta}_{jg}^{HB}\right)}}{\hat{\theta}_{jg}^{HB}},$$

where the posterior variance $Var\left(\hat{\theta}_{jg}^{HB}\right)$ is computed as

$$Var\left(\hat{\theta}_{jg}^{HB}\right) = \frac{\sum_{b=1}^{4,500}\left(\theta_{jg}^{(b)} - \hat{\theta}_{jg}^{HB}\right)^2}{4,500-1}.$$

Table 3-5 summarizes the distributions of the widths (the difference between the upper bound and the lower bound) of the credible intervals as well as the CVs for the 50 states and District of Columbia and the less than high school group in the United States for literacy proportion at or below Level 1.

Table 3-5. Distribution of credible interval widths and coefficients of variation for state-level model-based estimates for the less than high school group and for literacy proportion at or below Level 1: 2012/2014/2017 PIAAC

	Percentile				
Statistics for less than high school	20	40	60	80	Median
State estimates					
95 percent credible interval width (percent)	18.6	20.5	22.0	24.6	21.1
Coefficient of variation (percent)	9.8	11.2	12.8	17.0	11.8
Sampled state estimates					
95 percent credible interval width (percent)	18.3	20.4	22.0	24.5	21.0
Coefficient of variation (percent)	9.8	10.7	12.4	15.9	11.5
Nonsampled state estimates					
95 percent credible interval width (percent)	20.7	21.4	22.7	24.8	21.5
Coefficient of variation (percent)	12.7	13.0	15.6	17.7	13.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

The table shows that the median credible interval width is 21 percent for state estimates for the less than high school group with PIAAC sample cases and 21.5 percent without PIAAC sample cases. The CVs for the model-based estimates are mostly lower than 20 percent. Estimates with CVs of this magnitude are considered precise. But the users should note that for some states and other age and education groups, for other proportions (at Level 2 and at or above Level 3), the CVs can be large.

4. BENCHMARKING

In an earlier task, bivariate and univariate threefold HB models were used to produce the model-based model estimates for proportions (P1 and P3) and averages for all the U.S. counties. The model-based estimates were also aggregated to state level and nation level, and the results are displayed interactively in the U.S. PIAAC Skills Map (https://nces.ed.gov/surveys/piaac/skillsmap/).

The model estimates of proportions and averages that were produced in the areas defined by state and age and education groups, as described in section 3, can be aggregated across groups to state level, using state by group populations as the weight. In order to keep these aggregated state-level model estimates from age and education groups consistent with the state-level model-based estimates that were generated in the earlier task, benchmarking was implemented to the state-level model estimates for age and education groups. Since the state-level model estimates have a smaller mean squared error, benchmarking can be helpful to improve the precision of the state-level, model-based estimates for groups.

Benchmarking was applied after SAE model fitting but before producing the predicted values (see section 3.2.4 for details). The value $\theta_{jg}^{(b)}$ for each MCMC sample b was adjusted by a factor as follows:

$$\ddot{\theta}_{jg}^{(b)} = \theta_{jg}^{(b)} \frac{\theta_j^{(b)} T_j}{\sum_g \theta_{jg}^{(b)} T_{jg}},$$

where $\theta_j^{(b)}$ is the MCMC sample b from the posterior distribution of the Skills Map state-level proportions or averages, and T_{jg} and T_j are the population totals obtained from the 5-year ACS data, for state *j* and group *g*, and for state *j* across all groups, respectively.

After adjustment, the state-level MCMC samples, $\ddot{\theta}_{jg}^{(b)}$ for groups, if aggregated across groups denoted by g, would match the distribution of the state MCMC samples $\theta_j^{(b)}$. The benchmarked MCMC samples $\ddot{\theta}_{jg}^{(b)}$, in place of $\theta_{jg}^{(b)}$, were used to compute the posterior means $\hat{\theta}_{jg}^{HB}$ and precision measures $Var(\hat{\theta}_{jg}^{HB})$ in sections 3.2.4 and 3.2.5.

5. MODEL DIAGNOSTICS AND EVALUATION

An extensive model evaluation process occurred to ensure goodness of fit. Internal and external checks are illustrated in sections 5.1 and 5.2, respectively. Selected results are included for the literacy model for proportions at or below Level 1 for the less than high school group. The diagnostics and evaluations for other models are similar and can be found in the internal Supplemental Data Documentation.

5.1 Internal Model Validation

The internal model validation, conducted to check the model accuracy and robustness, was based on the set of 44 state-level estimates to which the models were fitted for each age and education group. Inference was conducted using 4,500 MCMC samples from the three chains combined. To assess the HB models, internal model validation checks were used as defined in the following sections.

5.1.1 Convergence and Mixing Diagnostics

Table 5-1 shows the mean and quartiles of the diagnostic statistics, including effective sample size, Gelman-Rubin $\hat{R}$ statistic, Monte Carlo standard error relative to the posterior standard deviation, autocorrelation, and cross-correlation, across all 182 monitored parameters. The results indicate that convergence and mixing of the three chains have been reached. For all the monitored parameters, the effective sample sizes are close to the total sample size (4,500 samples), the Monte Carlo standard errors are less than 10 percent of the posterior standard deviations, and the $\hat{R}$ factors are below 1.1. Autocorrelations within chains and cross-correlations among the monitored parameters are low. The cross-correlation is a matrix of dimension 182 by 182 with 1's on the diagonal entries.

Table 5-1. Convergence diagnostics for the Markov chain Monte Carlo (MCMC): 2012/2014/2017 PIAAC

Metric	$\hat{R}$	Effective sample size	Monte Carlo standard error/ posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9994	3912.719	0.0131	-0.0493	-0.0362	-0.9933
1st quantile	0.9997	4458.019	0.0147	-0.0109	-0.0128	-0.0307
Median	1.0000	4500.000	0.0149	-0.0034	-0.0048	-0.0005
Mean	1.0000	4544.418	0.0149	-0.0022	-0.0033	0.0161
3rd quantile	1.0003	4648.571	0.0150	0.0069	0.0067	0.0314
Maximum	1.0015	6436.789	0.0166	0.0463	0.0275	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.1.2 Multicollinearity Test

The variance inflation factor (VIF) was calculated to detect the multicollinearity among the model covariates. Table 5-2 shows that there is no indication of multicollinearity among the set of covariates that were used in the P1 model for the less than high school group. The variance inflation factors are all under 4 for other models.

Table 5-2. Variance inflation factors (VIFs)

Variables	VIF
Insurance	1.0300
Poverty	1.5109
Age_16_24	1.5345

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

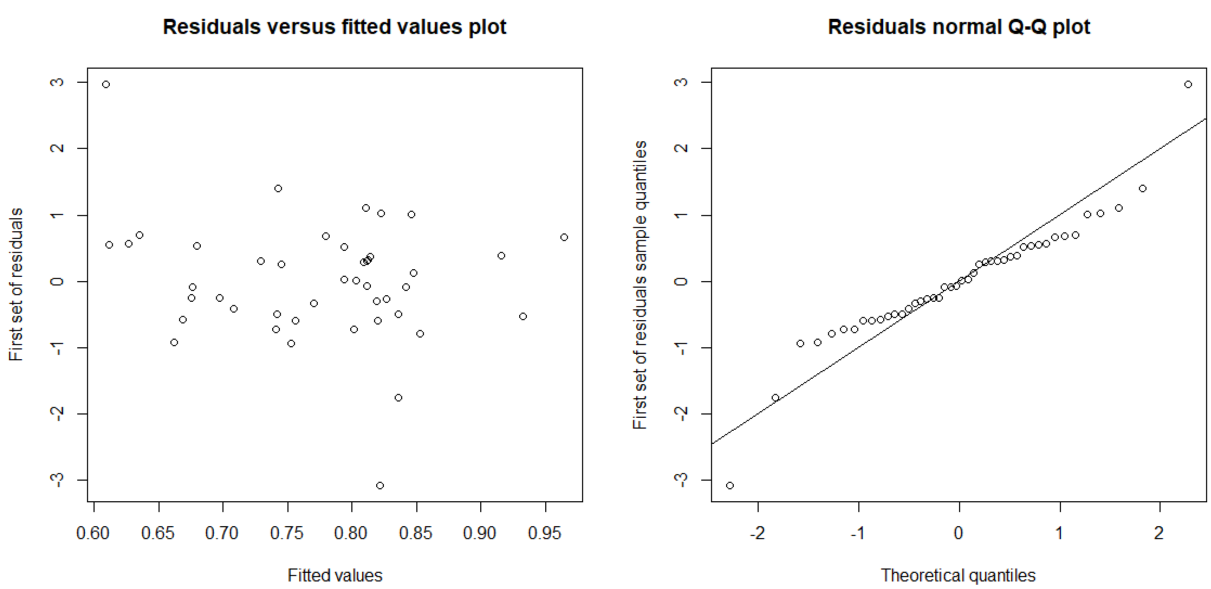
5.1.3 Residual Analysis

Naive normality checks using residuals indicate that there are no significant departures from normality, as shown in figures 5-1 and 5-2. Also, there is no significant pattern in the plot of residuals against the fitted values. For these checks, two sets of residuals are constructed:

$$r_{jg}^{Set1} = \frac{y_{jg} - x'_{jg}\hat{\beta}_g}{\sqrt{\hat{\sigma}_{v,g}^2 + \sigma_{jg}^2}}$$

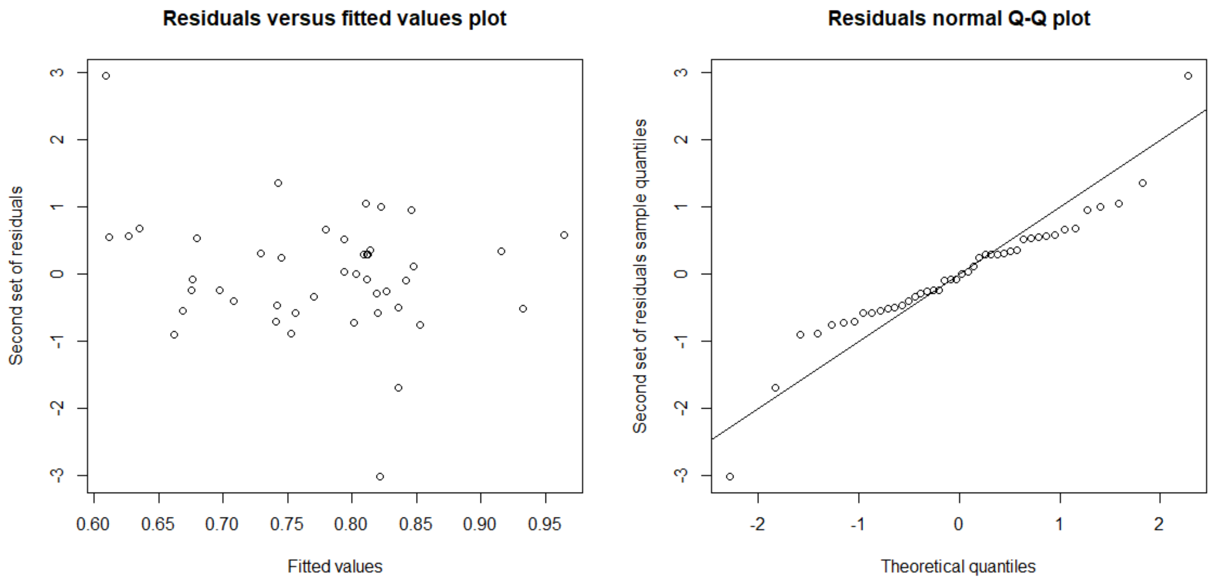
$$r_{jg}^{Set2} = \frac{y_{jg} - \hat{\theta}_{jg}}{\sqrt{\sigma_{jg}^2}}.$$

Figure 5-1. Residual plots for the first set of residuals



NOTE: The plots on the left-hand side illustrate the relationship between the residuals and the fitted values ($x'_{jg}\hat{\beta}_g$). The normal quantile plots are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 5-2. Residual plots for the second set of residuals (conditional on the random effect)



NOTE: The plots on the left-hand side illustrate the relationship between the residuals and the fitted values ($x'_{jg}\hat{\beta}_g$). The normal quantile plots are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

The residual plots with the fitted values being the x-axis indicate no significant departure from the homogeneous variance assumption. The two sets of residuals are very similar because the variance of the random effect is very small. Two model estimates have noticeably large residuals (i.e., greater than 2 in absolute value). They correspond to initial estimated state-level proportions of 0 and 1 for groups.

5.1.4 Posterior Predictive Checks

To assess the adequate fit of the model, posterior predictive checks were conducted. During the checks, statistics were compared based on their posterior predictive distribution to their corresponding values obtained using the original sample for a set of predefined statistics including identity, order indicator, mean deviation, unscaled residual constructed as deviation from the survey regression estimate, and scaled residual constructed as deviation from the survey regression estimate scaled by the posterior variance of the predictor using simulated values. Details about the construction of these statistics can be found in section 5 of the technical report.

The resulting posterior predictive *p* values are 0.504, 0.540, and 0.521 for the indicator and order and mean deviation statistics, respectively. Summary results for the posterior predictive statistics are provided in table 5-3 for literacy proportions at or below Level 1. The posterior predictive values for the ordered proportions indicate that values generated from the posterior predictive distribution are close to the sample values. The posterior predictive values for the indicator test statistics are close to 0.5, the deviations and the unscaled residuals are close to zero, and the scaled residuals range is not far from -1.96 to 1.96. Therefore, overall, there is no substantial indication for model lack of fit.

Internal diagnostic checks using the same set of methods and measures were done for the state-level models fitted to the other age and education groups, as well as for proportions at or above Level 3 and averages, and for numeracy. Detailed results can be found in the internal Supplemental Data Documentation. Although the diagnostic statistics vary across models, in general, there is no evidence of lack of fit for any of them.

Table 5-3. Summaries of posterior predictive statistics for literacy proportions at or below Level 1: 2012/2014/2017 PIAAC

Metric	Deviation	Unscaled residual	Scaled residual	Indicator	Order
Minimum	-0.0933	-0.9459	-2.8891	0.2273	0.0000
1st quantile	-0.0159	-0.0594	-0.4638	0.4545	0.3636
Median	0.0011	-0.0022	-0.0303	0.5000	0.5682
Mean	0.0015	-0.0085	-0.0110	0.5036	0.5395
3rd quantile	0.0188	0.0638	0.4911	0.5455	0.7273
Maximum	0.0958	0.4569	2.9133	0.7955	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

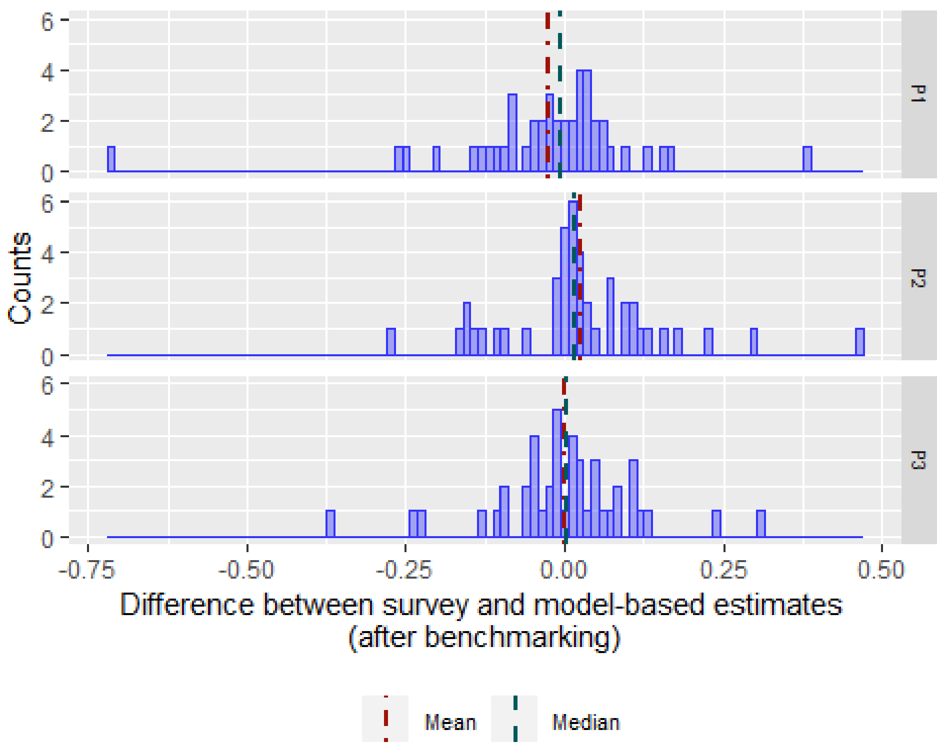
5.2 External Model Validation

External model validation results are illustrated thorough various graphs: (1) histograms of the difference between survey estimates and model-based estimates, (2) bubble plots of survey estimates versus model-based estimates, (3) shrinkage plots with arrows showing the direction from survey estimates to model-based estimates, (4) interval coverage plots, and (5) variance plots. The plots shown below are all based on the after-benchmarking, model-based estimates and only for the literacy proportion estimates for the less than high school education group. The before-benchmarking, model-based estimates behave similarly to the after-benchmarking, model-based estimates and are available in the internal Supplemental Data Documentation together with other proficiency estimates.

5.2.1 Histograms of Differences in Estimates

The differences between survey estimates and model-based estimates are shown in the histograms in figure 5-3. This plot shows the distribution of the difference as well as possible outliers. The means and medians of the differences are around zero. The majority of the differences are within 20 percentage points. The outliers in the plots show that a few model-based estimates could deviate from the survey estimates by about 20-75 percentage points.

Figure 5-3. Literacy proportions (less than high school) – Histograms of differences between survey estimates and model-based estimates: 2012/2014/2017 PIAAC

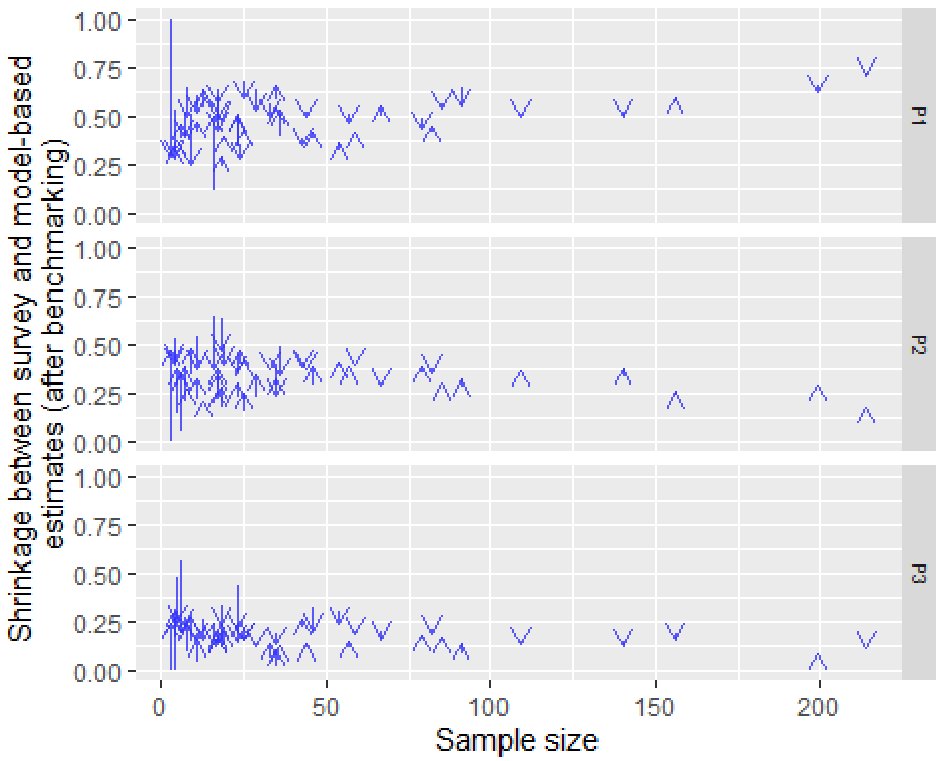


NOTE: The outlier with value around -0.7 for P1 is for a state with small sample size (n = 3) and has an extreme survey estimate (100 percent) for the literacy proportion at or below Level 1.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.2.2 Shrinkage Plots

Shrinkage plots show the direction from survey estimates to model-based estimates by sample size. The shrinkage can be observed in figure 5-4. As expected, the shrinkages are more significant in areas with smaller sample sizes than those in areas with larger sample sizes. The model-based estimates and the survey estimates become much more similar when the sample sizes are above 50.

Figure 5-4. Literacy proportion (less than high school) – Shrinkage plots of point estimates by sample size: 2012/2014/2017 PIAAC

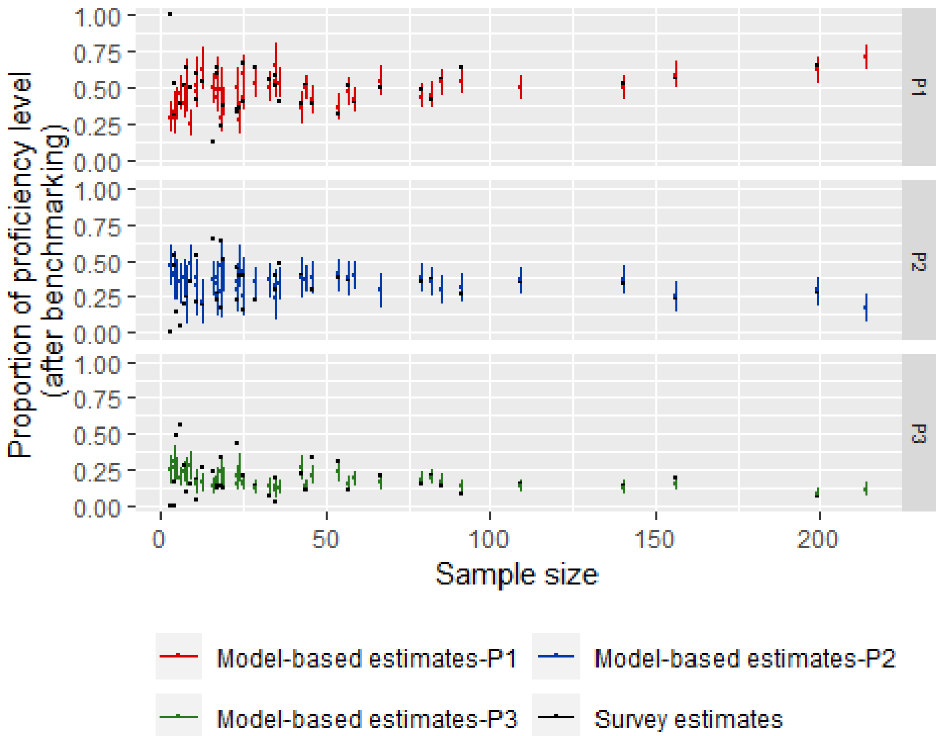


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.2.3 Interval Coverage Plots

The interval coverage plots in figure 5-5 show that, for a majority of the small areas, the 95 percent credible intervals generated from the models cover the survey estimates, especially for areas with large sample sizes. When the sample sizes are less than 50, sometimes the credible intervals for the model estimates do not cover the survey estimates. This is expected because the survey estimates contribute less to the model-based estimates if the survey estimates are derived from samples of smaller sizes (i.e., less reliable). Also, note that the credible intervals are narrower for small areas with larger sample sizes than with smaller sample sizes, as expected.

Figure 5-5. Literacy proportion (less than high school) – Indication of coverage by credible interval: 2012/2014/2017 PIAAC

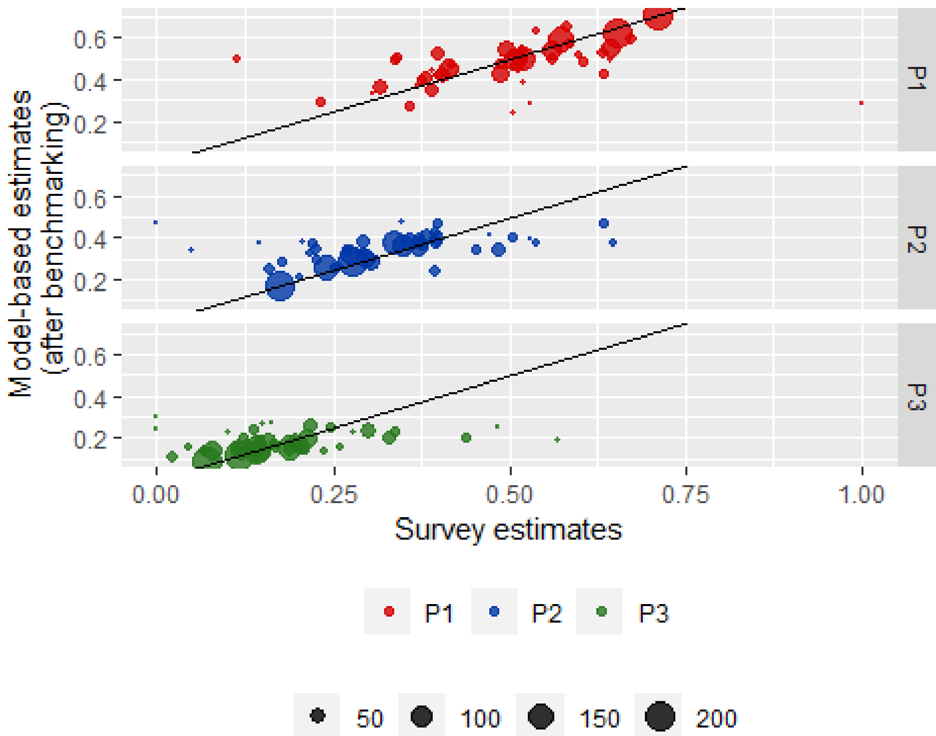


NOTE: The legend for the survey estimates are point estimates shown as the black dots in the figure.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.2.4 Bubble Plots of Survey Estimates and Model-Based Estimates

Figure 5-6 shows the survey estimates versus model-based estimates. The majority of the points are around the 45-degree line, indicating that the model-based estimates are close to the survey estimates. Larger bubbles (i.e., groups with larger sample sizes) have closer estimates than smaller bubbles. Some of the small bubbles, with the sizes of bubbles being proportional to the sample sizes in the small areas, are farther away from the 45-degree lines. This is as expected due to higher sampling errors for the survey estimates with small sample sizes.

Figure 5-6. Literacy proportion (less than high school) – Comparison between survey estimates and model-based estimates: 2012/2014/2017 PIAAC

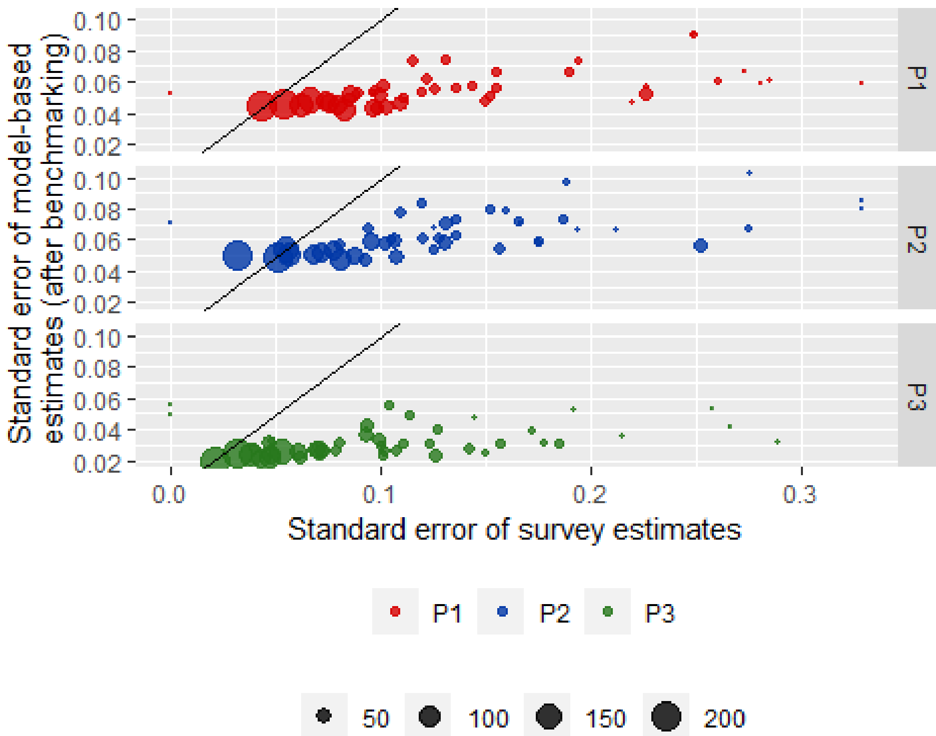


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.2.5 Survey Estimates and Model-Based Estimates Variances

Figure 5-7 shows the standard errors of the survey estimates and the posterior standard deviations from the small area model. While in general the resulting model-based estimates are similar to the survey estimates, users should keep in mind that the standard errors of proportions depend on the sizes of the estimated proportions. Therefore, if the model proportion is different from the survey proportion, the variance will in theory be different; therefore, the decrease in standard error is not necessarily an improvement due to the model. The plot shows that the model produces smaller posterior standard deviations than the survey standard errors, especially for areas of very small sample sizes.

Figure 5-7. Literacy proportion (less than high school) – Comparison between model standard errors and survey standard errors: 2012/2014/2017 PIAAC



SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6. SUMMARY

The PIAAC SAE process includes a statistical modeling approach that has taken into account significant enhancements made in SAE methodology. The statistical modeling approach was used to produce model-based estimates, which are available to the public on the Skills Map website at https://nces.ed.gov/surveys/piaac/skillsmap/. In 2020, the statistical modeling approach produced four state- and county-level estimates for adult literacy and numeracy proficiencies: an average score (on the PIAAC scale of 0 to 500) and the proportion of adults at or below Level 1, at Level 2, and at or above Level 3. The model-based estimates relied on the pooled 2012/2014/2017 PIAAC data as well as the ACS (2013-2017) data. The modeling depended on (a) PIAAC survey estimates; (b) area-level HB linear threefold models (bivariate for proportions, univariate for averages); and (c) seven covariates relating to educational attainment, poverty, race/ethnicity, health insurance coverage, and occupation (service industry).

In 2022, the statistical model approach was adapted to produce state-level, model-based estimates for six age groups and four education attainment groups. Like the state and county estimates, the model-based, state-level estimates for groups also relied on the combined 2012/2014/2017 PIAAC data and ACS (2013-2017) data. The modeling depended on (a) PIAAC survey estimates; (b) area-level HB linear univariate models with state-level random effects; and (c) four covariates for the age-group models (percentage of population with less than high school education, percentage of population with more than high school education, percentage of population below 100 percent of the poverty line, and percentage of population age 16 and over with service occupations) and three covariates for the education group models (percentage of civilian noninstitutionalized population who has no health insurance coverage, percentage of population below 100 percent of the poverty line, and percentage of population age 16-24).

The model was reduced from the state and county estimates model due to far fewer data points available for the state-level modeling for each group. Therefore, a selection process was developed and applied to select covariates from among the set of covariates in the state and county estimates model. A variety of methods was used to evaluate the fit of the HB models to the county estimates, including various methods of internal model validation as well as external model validation. The checks showed that the final models used were insensitive to different model assumptions and the measures indicated good fits to the data. A benchmarking approach was applied to align the aggregated, model-based estimates for groups to the state model-based estimates.

Overall, the state-level group model estimates are less precise than the state-level model estimates. For example, CVs for the state-level group model estimates for the less than high school group and for the proportion at or below Level 1 in literacy are generally of the order of 10 to 20 percent (see table 3-5), while the state predictions have a median CV of 8.1 percent (see table 5-5 in Krenzke et al. 2020). Estimates with CVs of this magnitude are considered to be precise for publication purposes. To a lesser extent, state-level estimates for groups from which some persons were sampled in the PIAAC 2012/2014/2017 combined household sample are more precise than estimates for states that had no persons sampled for an age or education group. With the positive diagnostics and evaluation results, the precision levels of the state-level model estimates for groups should give the data users confidence in using these model-based estimates.

In the PIAAC Skills Map, comparisons are available between small areas on the eight outcomes. That is, for literacy and numeracy, comparisons can be conducted on the proportion at or below Level 1, the proportion at Level 2, the proportion at Level 3 and above, and the average, for each group (age groups 16-24, 25-34, 35-44, 45-54, 55-64, 56-74; education groups: less than high school, high school, some college, and bachelor's degree or higher). The areas involved in the comparisons cover the following:

- State-to-nation
- State-to-state

Pairwise comparisons and multiple comparisons are also available as in the current Skills Map website at https://nces.ed.gov/surveys/piaac/skillsmap/. Details of the comparison methods can be found in the *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report* (https://nces.ed.gov/pubs2020/2020225.pdf).

REFERENCES

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6): 716-723.

Gelman, A., and Rubin, D. (1992). Inference From Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4): 457-472.

Hogan, J., Thornton, N., Diaz-Hoffmann, L., Mohadjer, L., Krenzke, T., Li, J., Van De Kerckhove, W., Yamamoto, K., and Khorramdel, L. (2016). U.S. Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014: Main Study and National Supplement Technical Report (NCES 2016-036REV). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Krenzke, T., Mohadjer, L., Li, J., Erciulescu, A., Fay, R., Ren, W., Van de Kerckhove, W., Li, L., Rao, J.N.K., Xie, H., and Provasnik, S. (2020). *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report* (NCES 2020-225). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Publishing Office. Retrieved on April 19, 2022 from https://nces.ed.gov/pubs2020/2020225.pdf.

Organization for Economic Cooperation and Development (OECD). (2016). *Technical Report of the Survey of Adult Skills (PIAAC)* (2nd ed.). Paris, France: OECD. Retrieved February 15, 2022, from https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf.

Rao, J.N.K. (2003). *Small Area Estimation* (Wiley Series in Survey Methodology). New York: Wiley.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation* (Wiley Series in Survey Methodology) (2nd ed.). Hoboken, New Jersey: Wiley.

Wolter, K.M. (2007). Taylor Series Methods and Generalized Variance Functions. In *Introduction to Variance Estimation*. Statistics for Social and Behavioral Sciences. New York: Springer.