Proposal for a NAEP SES Index: Background, Findings, and Remaining Challenges

AIR - NAEP Working Paper 2023-02

Markus Broer, Yifan Bai, Rebecca Shipan

JUNE 2023

The research contained in this working paper was commissioned by the National Center for Education Statistics (NCES). It was conducted by the American Institutes for Research (AIR) in the framework of the Education Statistics Services Institute Network (ESSIN) Task Order 14: Assessment Division Support (Contract No. ED-IES-12-D-0002/0004) which supports NCES with expert advice and technical assistance on issues related to the National Assessment of Educational Progress (NAEP) and the working paper was completed under a follow-on contract (#91990022C0053). AIR is responsible for any error that this report may contain. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Advancing Evidence. Improving Lives.

AIR

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

NCES

The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education in the U.S. and other nations. NCES is located within the U.S. Department of Education and the Institute of Education Sciences. NCES fulfills a Congressional mandate to collect, collate, analyze, and report complete statistics on the condition of American education; conduct and publish reports; and review and report on education activities internationally.

NCES support contracts

The analysis for this working paper were completed under Task Order 14 for which AIR is the prime contractor under the Education Statistics Services Institute Network (ESSIN). ESSIN is a network of companies that provided the National Center for Education Statistics (NCES) with expert advice and technical assistance, for example in areas such as statistical methodology; research, analysis and reporting; and survey development. This AIR-NAEP working paper is based on research conducted under the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14. The working paper itself was completed under a follow-on contract (#91990022C0053).

William Tirre, a Program Director in the NCES Assessment Division, oversees the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14 and its follow-on contract referenced above.

Suggested citation:

Broer, M., Bai, Y., and Shipan, R. (2023). *Proposal for a NAEP SES Index: Background, Findings, and Remaining Challenges.* [AIR-NAEP Working Paper #2023-02]. Washington, DC: American Institutes for Research.

For inquiries, contact:

Markus Broer, Project Director for Research under NAEP Support Contracts
Email: mbroer@air.org

Mary Ann Fox, AIR Vice President and Project Director of NAEP Support Contracts
Email: mafox@air.org

Executive Summary

The National Assessment of Educational Progress (NAEP) is required by law to report on subgroup performance differences, such as those by race/ethnicity, gender, and student family socioeconomic status (SES). Since 2003, the most prominently used proxy variable for SES in NAEP has been National School Lunch Program (NSLP) eligibility. However, concerns have been raised that this indicator may become less valid over time, and as a binary variable, it has clear limitations both for reporting as well as for being used as an SES control variable in education research.

National Center for Education Statistics (NCES) convened an SES panel that discussed challenges and recommendations for SES measurement in NAEP. The panel's major recommendations included developing a composite measure of SES with a core measure based on the "big three" SES components (parental educational attainment, parental occupational status, and household income) that could be augmented with a school or neighborhood SES component. To address these recommendations, NCES developed and evaluated new SES proxy variables, while also encouraging research aimed at exploring the possibility of forming a NAEP SES index with core components that have long been collected by NAEP and validating the utility of such an index.

In this report, we propose a composite measure of SES for NAEP at grades 8 and 12 using four components that have been collected since 2003: (1) the number of books at the student's home; (2) the student's NSLP eligibility status; (3) the percentage of students eligible for NSLP at the school the student is attending; and (4) the highest level of education of either parent. At grade 4, where information on parental education is not collected from students, the index has only three components. The addition of the school SES component is a response to the NCES SES panel's recommendation to consider adding such a component to an SES composite measure.

To evaluate the NAEP SES index's performance and utility, NCES and the American Institutes for Research (AIR) developed three criteria:

- How well does the index explain NAEP performance?
- How well does the index account for racial/ethnic achievement gaps in NAEP?
- Does the index function similarly for major racial/ethnic subgroups?

The results show that the proposed NAEP SES index performs well against the established criteria:

- It explains a relatively large amount of performance variation in NAEP compared to using NSLP eligibility alone as a proxy variable. It also performs better than similar SES indices used in other large-scale surveys administered in the United States: the Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA), and the High School Longitudinal Study of 2009 (HSLS:09).

- It accounts for a larger share of the achievement gap between White students and Black, Hispanic, and Native American/Alaskan Native students compared to using NSLP eligibility alone as a proxy variable.
- It performs similarly within each major racial/ethnic subgroup; that is, higher NAEP SES index scores are associated with higher average NAEP scores within each racial/ethnic group.

While there are still challenges that need to be addressed before operational implementation, we believe that the proposed NAEP SES index would be a very useful addition to helping NCES and the NAEP program in its continuous efforts to further contextualize NAEP performance and understand the circumstances associated with varying educational performance.[1]

[1] **Acknowledgements**: The authors would like to thank NCES for funding and guiding this work; Bill Ward, who leads the SES Working Group at NCES; William Tirre, who oversees NAEP research conducted by AIR including all iterations of this work; our colleague from AIR, George Bohrnstedt, who has provided invaluable mentorship and guidance throughout this project; Qingshu Xie from MacroSys who was the lead analyst of exploratory work related to creating a NAEP SES index; and everybody from NCES, ETS, and Westat who have provided thoughtful critique and comments at various stages.

Contents

List of Tables

List of Figures

Introduction

The role of family socioeconomic status (SES) in students' educational outcomes has been a critical research issue in the education field for a long time. For most children, home is their first learning place and thus has a long-term influence on their cognitive and noncognitive development (Choi and Byun 2022). Substantial research has shown that students from higher-SES family backgrounds tend to perform better at school (Liu, Peng, and Luo 2020; Sirin 2005) compared to students from lower SES backgrounds who face a myriad of obstacles that can affect their learning (Marks and Pokropek 2019; Lareau 2011). However, it has been a challenge to reach a consensus about the conceptual definition of SES and the operational construct of an SES measure in empirical research.

The National Assessment of Educational Progress (NAEP) is required by law to report on subgroup performance differences, such as those by race/ethnicity, gender, and family SES background. Since 2003, the most prominently used NAEP proxy variable of student SES has been National School Lunch Program (NSLP) eligibility. However, there are concerns that this indicator may become less valid over time (Cowan et al. 2012), especially since nationwide implementation of the Community Eligibility Provision (CEP)[1] in the 2014–15 school year under the Healthy, Hunger-Free Kids Act of 2010.

Domina et al. (2018) found that data on NSLP eligibility may better capture students' educational disadvantage than IRS-reported annual household income data, and recent work with NAEP data has also pointed to the continued relevance of NSLP status as an SES proxy (Broer, Chen, and Xie 2019; Xie and Broer 2022a). However, NSLP eligibility, as a binary variable, has clear limitations both for reporting and for being used as an SES control variable in education research.

In 2012, an expert panel was convened by the National Center for Education Statistics (NCES), which recommended developing an SES index for NAEP (Cowan et al. 2012). The panel voiced concern about the use of NSLP eligibility as an SES proxy and made some recommendations to improve the measurement and reporting of SES, including the construction of a core SES composite score (based on students' family resources, parental educational attainment, and parental occupational status) and an expanded SES measure that could include neighborhood and school SES.

Under NCES guidance, researchers from the American Institutes for Research (AIR), the Educational Testing Service (ETS), and Westat have worked on various aspects of this effort, from item development and data collection methods to research and validation. The NAEP SES index proposed here is the culmination of research efforts dating back to 2015 (starting with

[1] The Community Eligibility Provision (CEP) is a non-pricing meal service option for schools and school districts in low-income areas. More details at https://www.fns.usda.gov/cn/community-eligibility-provision.

unpublished research and presentations to NCES), with the first results for a version of the proposed SES index presented at the 2017 American Educational Research Association (AERA) conference (Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017).

The proposed SES index will enable a finer-grained analysis of how SES is associated with NAEP performance. Such an analysis could even be done retroactively, given that the variables used in the index have been collected since 2003. A variation of the index has already been used in an AIR-NAEP research report whose goal was to understand the development of achievement gaps between low- and high-SES students between 2003 and 2017, for the nation as well as across states (Bai, Straus, and Broer 2021).

The NAEP Alliance[2] has proposed using this composite SES index as a starting point to support NAEP reporting. NCES has expressed special interest in understanding potentially differentiated COVID-19-induced learning lags and subsequent recovery for students from different SES backgrounds.

The latest pre-pandemic NAEP data from 2019 is used in this report as a starting point to familiarize stakeholders with the progress made in developing a NAEP SES index, explain how the index functions, illustrate how such an index could be used for reporting, and discuss various challenges associated with operational implementation. It is not intended as an academic paper or technical documentation and its scope does not allow for a thorough discussion of all previous related efforts or in-depth theoretical and measurement considerations.

We believe that the results on the performance of the proposed NAEP SES index with respect to the established evaluation criteria are positive and we hope that the information contained in this report will prove to be useful for furthering the next steps toward an operational implementation of an SES index in NAEP.

Background

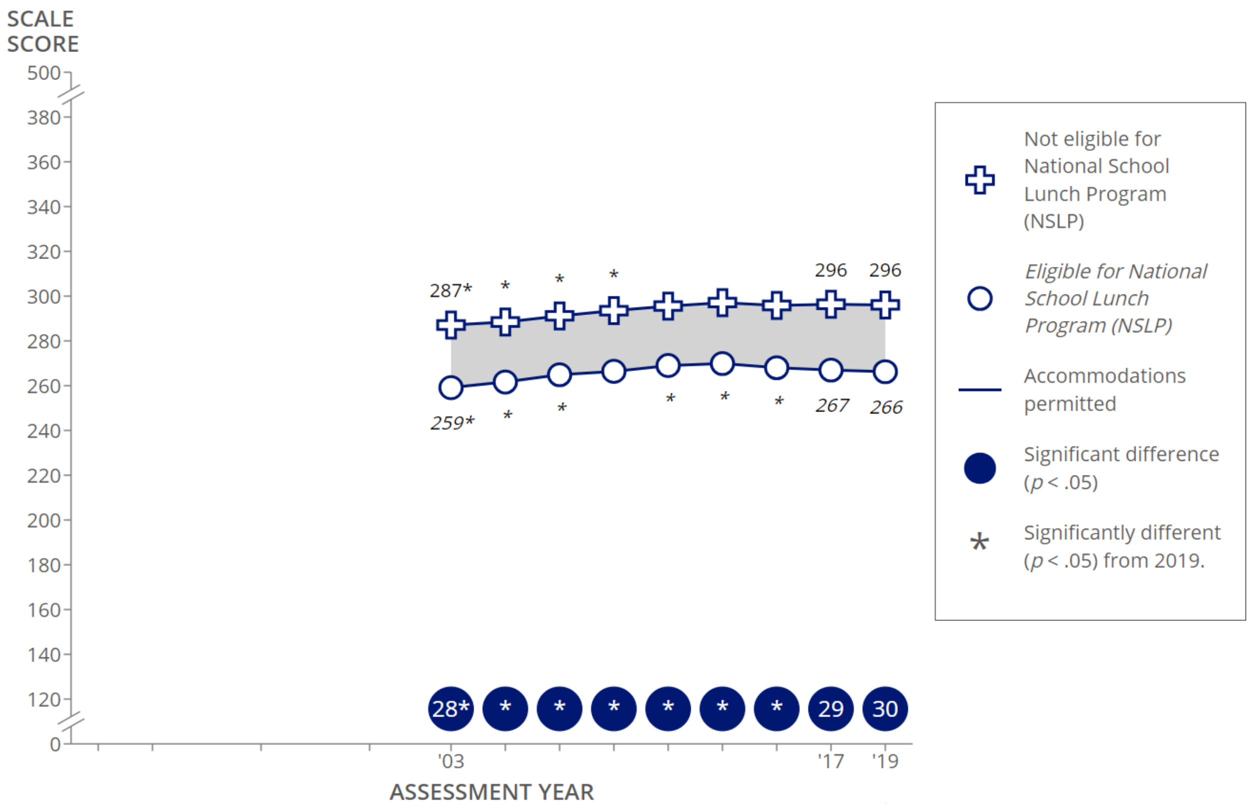
Current SES reporting in NAEP

As noted in the Introduction, the main SES reporting variable in NAEP has been students' NSLP status. Setting aside students for whom no information is available on this variable, the score gap between NSLP-eligible students (lower SES) and non-eligible students (higher SES) can be reported for the current year and compared to gaps in previous NAEP administrations. Figure 1, using data from the Nation's Report Card of 2019, shows that scores in grade 8 mathematics rose for both NSLP-eligible students and non-eligible students from 2003 to 2019. However, despite the many intervening changes in that time period, the score gap between the two

[2] The NAEP Alliance network includes NAEP assessment operation contractors, such as ETS and Westat.

groups increased only slightly (28 points in favor of non-eligible students in 2003 vs. 30 points in 2019). In addition, the rise in schoolwide lunch programs has led to concerns that NSLP as an income proxy may have become less effective at distinguishing between students from different income levels.

Figure 1. Trend in average scale scores and score gaps for grade 8 NAEP mathematics, by National School Lunch Program (NSLP) eligibility: Various years, 2003–2019

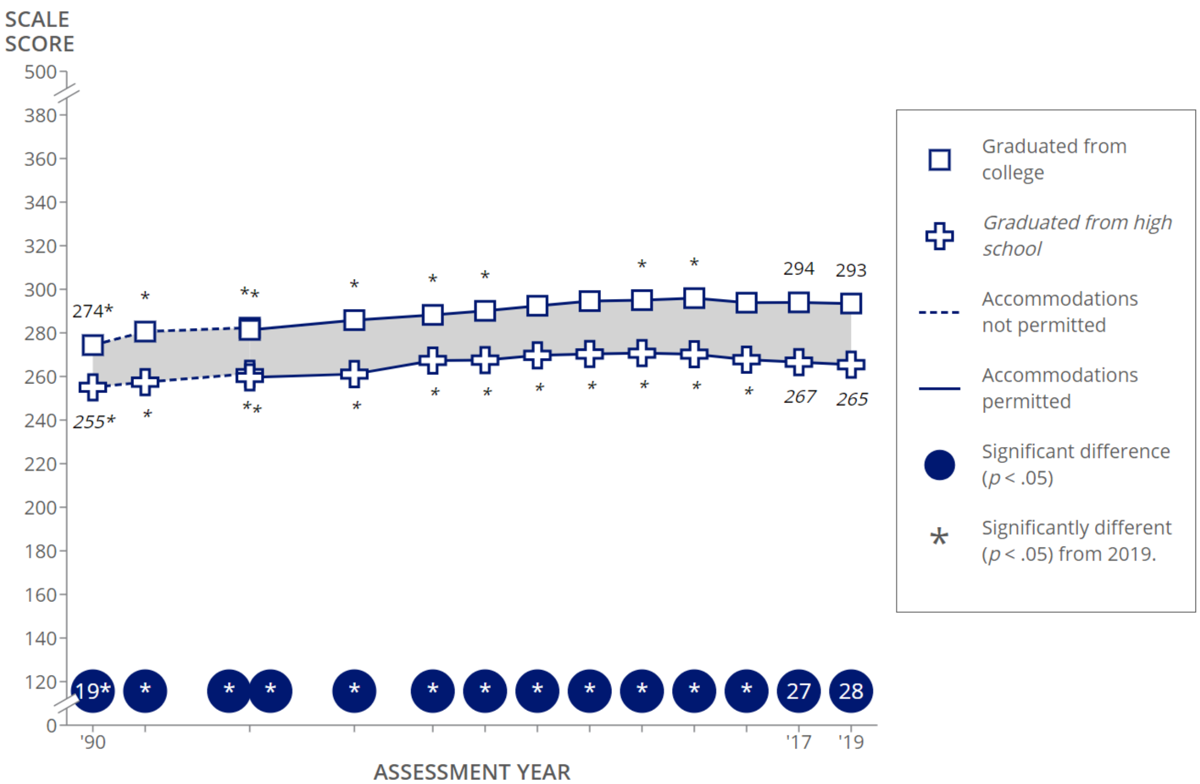


* $p < .05$. Significantly different from 2019.
NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. Differences were calculated using unrounded values. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, and 2019 Mathematics Assessments. https://www.nationsreportcard.gov/mathematics/nation/groups?grade=8#nation-gaps-gaps

Another variable used in reporting SES is "highest level of parental education of either parent." However, reporting SES using the parental education variable alone also has limitations. First, it is not available across all grade levels; this variable is not collected in the grade 4 student questionnaire, as research has shown that fourth-graders are not able to report reliable information about their parents' educational attainment (Jewsbury et al. 2016; Musu-Gillette 2016). Second, even in grades 8 and 12, the missing rate for the parental education variable is relatively high due to various factors: some states have opted out of including this question in the questionnaire entirely, some students skip the question, and other students answer, "I don't know." Jewsbury et al. (2018) suggested that the missing mechanisms could be varied, and that the parental education variable is likely not missing at random. Therefore, listwise deletion is not an option for reporting, but including the missing category also means that the percentages for the valid responses are likely different from the actual distribution for that variable.

Figure 2, based on data from the Nation's Report Card of 2019, uses the "highest level of parental education of either parent" variable to present a binary comparison of the gap in average NAEP performance between students who reported that their parent(s) graduated from college and those who reported that their parent(s) graduated from high school. Again, scores for both groups rose between 1990 and 2019, but in this case the scores for the higher-SES group (i.e., students whose parents graduated from college) rose at a faster rate, resulting in an increase in the score gap between the two groups over time (19 points in 1990 vs. 28 points in 2019).

Figure 2. Trend in average scale scores and score gaps for grade 8 NAEP mathematics, by highest level of parental education: Various years, 1990–2019



* $p < .05$. Significantly different from 2019.
NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. Accommodations were not permitted in NAEP mathematics assessments prior to 1996 at the national level for grade 8. Differences were calculated using unrounded values. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2000, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, and 2019 Mathematics Assessments. https://www.nationsreportcard.gov/mathematics/nation/groups?grade=8#nation-gaps-gaps

The two other variables proposed for inclusion in the SES index—school SES and books at home—are less often used in reporting and research, even though both have relatively high correlations with NAEP performance. Xie and Broer (2022b) found that the percentage of students eligible for NSLP at the school a student is attending had a correlation coefficient with student performance of $r = 0.36$ using 2017 NAEP mathematics grade 8 public school data. In comparison, the School Neighborhood Poverty Estimates from the NCES Education Demographic and Geographic Estimates (EDGE)[3] program had a correlation coefficient of $r = 0.29$. For books at home, the correlation with NAEP performance had a coefficient of $r = 0.38$.

Nevertheless, reporting by individual SES-related variables has long been seen as having limited utility.

Recommendations from NCES panel on SES

In 2012, NCES convened an expert panel to make recommendations on SES reporting options and address the shortcomings of reporting based on individual variables, especially NSLP status.[4]

Concerns about using NSLP eligibility as the main SES proxy variable for reporting have been raised for a long time, including concerns about the fact that it captures only one aspect of SES (related to income), about underestimation of NSLP at grade 12, about the essentially binary nature of the variable, and about how schoolwide lunch programs may impact the quality of NSLP data and how they are collected and used in NAEP (Cowan et al. 2012).

The expert panel, which represented various fields (including economics, education, statistics, human development, and sociology), was tasked with providing recommendations concerning the measurement of SES in NAEP. The concluding report, *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress* (Cowan et al. 2012), offered several recommendations that will briefly be described here (refer to the full report for additional details).

The panel broadly defined SES as ***one's access to financial, social, cultural, and human capital resources***, while also acknowledging the "big three" SES components used in traditional sociological definitions: parental educational attainment, parental occupational status, and household or family income.

---

[3] The NCES Education Demographic and Geographic Estimates (EDGE) program publishes a School Neighborhood Poverty Estimates based on household economic data from the Census Bureau's American Community Survey (ACS) and the geographic locations of public schools. This is then used to estimate the income-to-poverty ratio for neighborhoods around school buildings.

[4] NAGB also convened an expert panel in October 2011. One recommendation was as follows: "The development and use of improved measures of socioeconomic status (SES) will be accelerated, including further exploration of an SES index for NAEP reporting."

The panel made **four key recommendations** to NCES to improve the measurement and reporting of SES:

(1) Develop indicators or proxies of the "big three" SES components to form the core SES measure used in reporting;
(2) Add school or neighborhood SES to an expanded measure of SES;
(3) Attempt to develop a composite measure (index) of SES; and
(4) Due to perceived data quality issues with NSLP status and student reports of parental educational attainment and occupational status, especially in grade 4, explore linking to Census data on SES components.

NCES initiatives in response to the SES panel's recommendations

The recommendations of the SES panel have been the basis for NCES's various research and development efforts on SES measurement. To oversee this work, a NAEP SES working group was formed, led by NCES Senior Research Scientist William Ward and including NCES staff and contractors' representatives. This group continues to meet regularly to discuss plans and to review empirical efforts in this area.

In response to recommendation #1 on focusing on measuring "the big three" SES components, NCES has dedicated many resources to the piloting of new questionnaire items and the reevaluation of existing ones.

Since 2013, the NAEP student core questionnaire has included questions on household structure and access to resources for all grade levels. The NAEP program has also removed some questions about home resources because they could be seen as intrusive and/or had weak relationships to student achievement. Moreover, research by both AIR and ETS has shown that household possessions generally do not relate well to other SES components nor to NAEP performance (especially after accounting for the relationship between performance and other SES components) and would therefore, among other reasons, not lend themselves well to inclusion in a composite index (see Jewsbury et al. 2018).

Questionnaire development has also focused on finding a proxy variable for one of the "big three" SES components—occupational prestige—for which no variable has been available in NAEP. Based on an analysis of pilot data, reporting of parental occupation categories was not deemed reliable for both grade 4 and grade 8 students (Jewsbury et al. 2016) and thus parental occupation was not included in 2017 operational questionnaires.

In 2017, a new question on parents' occupational status was added ("Does your mother work? Does your father work?"). However, investigations by both AIR and ETS (Jewsbury et al. 2018) showed that occupational status, as measured by these items, did not relate well to other SES components and did not contribute further to the explanation of NAEP performance; that is, it was shown that the inclusion of this proxy variable in an SES index would not improve the SES index's relationship to NAEP performance.

NAEP has also experimented with the wording of questions on parental education. Focus groups of parents and students were formed to evaluate the questions, and empirical research was conducted using existing data. For example, an ETS research study evaluated data from an overlap sample between fourth-graders who participated in both NAEP and the Early Childhood Longitudinal Study of 2011 (ECLS-K:2011) to assess if students can reliably report parental education by comparing answers from students and their parents. The conclusion was that parental education could not be reliably collected for grade 4 students (Jewsbury et al. 2016).

In summary, despite several survey development efforts since the SES panel report, no new questionnaire item developed since 2012 has been shown to improve the accuracy of a NAEP SES index. In effect, the existing items that have been collected since 2003 still appear to be the best option for forming an SES index.

In response to recommendation #2 on considering measures of school and neighborhood SES status, two developments are noteworthy. First, the NCES EDGE program used geospatial methods to develop School Neighborhood Poverty Estimates for all public schools based on household economic data from the Census Bureau's American Community Survey (ACS) and the geographic locations of public schools (Geverdt 2018).

Second, AIR conducted analyses showing the usefulness of including the percentage of NSLP-eligible students in the school as a component in an SES index (Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017).

Later, AIR also compared the two school SES (or school poverty) measures—the percentage of NSLP-eligible students in a school and the NCES EDGE program's School Neighborhood Poverty Estimates—in terms of (a) how they each relate to other SES components as stand-alone measures, as well as how they relate to NAEP performance; and (b) how an SES index performs when one or the other is used as the school poverty component of the index. Both measures worked comparatively well as a component of the proposed SES index. While "percentage of NSLP-eligible students in a school" worked better on some measures and is the component used in this report, the NCES EDGE measure on school poverty also worked well and therefore can be considered a viable alternative (Xie and Broer 2022b).

In response to recommendation #3 that NCES attempt to develop a composite SES measure, preliminary research to combine available SES components collected by NAEP into an SES index (as well as evaluating the inclusion of new variables as they became available) has been ongoing for several years through various efforts. AIR's preliminary internal analysis about a possible NAEP SES index started in 2015 (internal briefs; Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017). Westat also developed and presented an alternative SES index focused on household wealth (Tang 2017). ETS, in an internal document, presented psychometric considerations and evaluations related to the development of an SES index and examined an alternative SES index, in addition to the one proposed here and the one proposed by Westat (Jewsbury et al. 2018).

The research, considerations, and challenges mentioned above have been presented and discussed in the NAEP SES Working Group as well as in other NAEP committees: the NAEP Questionnaire Standing Committee (in 2017), the NAEP Design and Analysis Committee (in 2018), and the NAEP Validity Studies Panel (in 2019). Moreover, a presentation was made to the NAGB Reporting and Dissemination Committee in 2021. The results of these efforts have also been presented at conferences and workshops (e.g., AERA, STATS-DC, and the NAEP Assessment Literacy Workshop). Feedback from these presentations has been generally positive and the NAEP Alliance has proposed trialing the use of this composite SES index to support NAEP reporting.

In response to recommendation #4 on linking to Census data given the data quality concerns about the NSLP and other variables, Xie (2019) carried out a comparison of NSLP and ACS data. The investigation found that NAEP-reported NSLP eligibility rates are higher than estimates of NSLP eligibility based purely on income information from ACS at all three grades, likely reflecting the intended effects of specific policies aimed at increasing access to free lunch for students who attend schools in areas of concentrated poverty. The investigation also showed that ACS data reflect the decline in NSLP eligibility rates seen at higher grade levels in NAEP. This is an important finding suggesting that the decline in NSLP eligibility rates with the rise in grade levels is fundamentally determined by the increase in family income associated with the increase in student age, not mainly due to decisions by students or their parents that would reduce their NSLP eligibility rates.

NAEP also collected ZIP code information from students starting in 2013, an approach that would allow a linkage to ACS data. However, the information had a high missing rate and was deemed unreliable. The ZIP code question was discontinued in 2019 and is no longer part of student questionnaires.

Unrelated to the Census data, other ways of examining the relationship between NAEP SES proxies and external data have been explored. For example, Ogut, Bohrnstedt, and Broer (2016) investigated the overlap sample of NAEP grade 12 and the High School Longitudinal Study of 2009 (HSLS:09) to understand how NAEP SES proxies[5] are related to parent reports of the "big three" SES components collected in HSLS:09. The results indicated that NAEP SES proxies could best predict HSLS:09's parental education component (R^2 = 0.48), followed by family income (R^2 = 0.31) and occupational prestige (R^2 = 0.23). Additionally, the best individual NAEP predictor of family income in HSLS:09 was whether a student was eligible for free or reduced-price lunch; the best predictor of parental education in HSLS:09 was—unsurprisingly—parental education in NAEP; and the best predictor of occupational prestige in HSLS:09 was student reports on parental education in NAEP.

[5] Number of books at home, parental education, NSLP eligibility, and household possessions (computer, Internet, dryer, dishwasher, multi-bathroom, own bedroom)

This effort led to a first attempt to construct a NAEP SES index based on the relationships found in the overlap sample between the HSLS:09 SES index and NAEP SES proxies (Ogut, Bohrnstedt, and Broer 2016). However, further analysis of the NAEP overlap sample for which the HSLS:09 SES index was also available revealed that the HSLS:09 index did not perform as well as expected, despite being composed of the "big three" SES components and the data being provided directly by parents (Broer, Xie, and Bohrnstedt 2017). This realization was the basis for exploring whether a NAEP SES index based on measures currently collected by NAEP might perform at least as well as, or better than, the HSLS:09 SES index, based on criteria that will be laid out below. For more detail about the HSLS:09 SES index, please refer to Appendix G.

The same overlap sample was also used to compare student reports on parental education for grade 12 with those from parents in HSLS:09. This comparison showed some divergence but generally good alignment (Ikoma, Bai, and Broer 2017). As mentioned in the discussion of the response to recommendation #1, researchers have used the overlap sample between NAEP and the ECLS-K:2011 to investigate the same issue for grade 4, concluding that student reports on parental education were not reliable (Jewsbury et al. 2016). ETS also investigated how NSLP status was related to income information provided by parents in ECLS-K:2011 using the same overlap sample and found high consistency between NSLP eligibility information and parent-reported household income (Jewsbury et al. 2016).

Regarding the perceived quality issues noted in recommendation #4 about the use of NSLP as a measure, especially related to concerns about its continued validity after widespread introduction of the Community Eligibility Provision (CEP), investigations by Broer, Chen, and Xie (2019) and Xie and Broer (2022a) have shown its continued usefulness as an income proxy. Despite the sharp increase in the percentage of students who attend schools that operate schoolwide lunch programs, the percentage of students coded as NSLP-eligible in NAEP has barely changed since the widespread introduction of CEP. Perhaps concerns have been based on the misconception that, since all students receive free lunch in schools that administer schoolwide free lunch, all students would therefore be categorized as NSLP-eligible in NAEP. But this is not the case in most instances, and only in a very small percentage of schools is such differentiating information not available.[6] Instead, NAEP follows states' definitions of "economic disadvantage," and students who are identified as being economically disadvantaged through various state criteria are then categorized as NSLP-eligible in NAEP. Students not meeting these state definitions are coded as non-eligible.

[6] Such schools are likely to be high-poverty schools (in which case most students in the school would be NSLP eligible).

Based on what has been discussed in this section, we have reached the following conclusions regarding the recommendations in the NCES SES panel report:

- We endorse the recommendation about forming a composite measure of SES (recommendation #3) and have included a school poverty component in the proposed composite measure of SES (recommendation #2).
- NSLP eligibility (or "economic disadvantage" status), while limited as a sole SES proxy, should continue to be a component of an SES composite for the foreseeable future (in response to recommendation #4). It is the only "objective" individual student measure related to income among large-scale surveys in that it is reported to NAEP based on school records rather than collected through student or parent reports in questionnaires. In addition, it has a strong relationship with NAEP performance.
- There is no viable proxy for occupational prestige, which means that the "big three" SES components cannot all be represented in a NAEP SES index (see recommendation #1), but this should not preclude efforts to construct a useful SES index for NAEP. The *Contextual Information Framework for the National Assessment of Educational Progress* (NAGB 2013) also stated, pragmatically, that "although NAEP may never be able to produce a full composite of SES, based on family income, education, and occupation, efforts should be accelerated to develop and use improved measures of socioeconomic status, including an SES index." Related to this point, our analysis of the occupational prestige component of the HSLS:09 SES index showed that it did not have as strong a relationship with students' grade 9 algebra scores ($r = 0.15$) as the other two components (parental education: $r = 0.30$, income: $r = 0.31$). Furthermore, occupational prestige did not explain unique performance variance beyond what was explained by the income and parental education components (Broer and Xie 2018). NAEP is not the only survey that does not include an occupation or occupational prestige measure. For example, the Trends in International Mathematics and Science Study (TIMSS) home educational resources (HER) index does not include one either (for more detail about the TIMSS HER index, see Appendix G).
- "Number of books at home" as a home possession item was initially thought to be an income proxy, but investigation by AIR (Ogut, Bohrnstedt, and Broer 2016) and ETS (Jewsbury et al. 2018) showed that it is most closely related to parental education. Given the broader definition of SES used by the panel, the "books at home" variable could also be conceived of as being part of "cultural capital." Furthermore, "books at home" has one of the strongest relationships with NAEP performance and should have a prominent place in an SES index (i.e., not merely one of many home possession items, which would decrease its importance).

A short description of approaches to SES index construction and indices used by other large-scale assessments

SES is seen as a formative, rather than a reflective, construct, meaning that the different levels of the chosen SES indicators are not determined by one underlying SES construct, but rather that the construct of SES emerges when the different SES components are combined (Jewsbury et al. 2016). Cowan et al. (2012) noted that while latent variables in reflective measurement are required to covary to a high degree, this is not a precondition for SES (or for other formative constructs).

There are different approaches to forming a composite index and weighting the different components, including the use of summative scores, scores based on principal component analysis, and weighting based on relationships to an outcome variable (in our case, NAEP performance). The “empirical” weighting methods employed by principal component analysis are more in line with a reflective measurement model. Summative scores based on the different levels of the component variables, as used in the current study, can be seen as arbitrary, but Cowan et al. (2012) noted that an advantage of their use is the ease with which the rules can be communicated to data users.

SES is also often seen as a multidimensional construct rather than a unidimensional one (see Jewsbury et al. 2016 for NAEP or Eriksson et al. 2021 for PISA), which presents its own challenges, e.g., that any component items are not simply interchangeable with any other items; replacement items would need to come from the same factor as the original item and function in a similar fashion. While the SES components available in NAEP (especially those that would have the strongest relationship with NAEP performance) are not unidimensional, they are still correlated. Combining them into a composite measure rather than using several separate indicators also has practical utility for reporting, offering a clearer and finer-grained analysis of the relationship between NAEP performance and SES than reporting by different single SES components would. Additionally, a composite SES index has practical utility as a control variable in education research, whereas the use of several individual SES components as control variables would create problems in analysis due to their correlated nature (Avvisati 2020).

TIMSS has used a categorical approach in the past, reporting by three categories based on the specific combination of the three SES components in its HER index (parental education, books in the home, and home study resources), but is now using an IRT-based measure. The Program for International Student Assessment (PISA) standardizes its SES index across countries and weights the three components equally (parental education, parental occupation status, and home possessions, the latter of which has frequent changes in its subitems). Alone among the assessments reviewed here, the HSLS:09 SES index components are based solely on the “big three” SES components, which are standardized before forming a composite. For more detail about the SES indices in these surveys, please refer to Appendix G.

Jewsbury et al. (2018) evaluated different weighting methods with NAEP 2017 grade 8 mathematics data: a sum-score-based approach, using the first component in a principal component analysis, and a clustered first principal component analysis approach. The results were very highly correlated, which was expected given prior research showing that the quality of the items included in a sum score is more important than how they are weighted (Wainer 1976, as quoted in Jewsbury et al. 2016). Based on a review of the literature on differential weighting, Stanley and Wang (1970) concluded that in practice the gain realized from using differential weighting is not worth the effort. Also, Broer, Bai, and Fonseca (2019) showed that for TIMSS data, the results were very similar for an SES index they constructed using a sum score method and for an index constructed using the weights from the first principal component.

For the NAEP SES index proposed here, we have opted to form a composite index based on a sum score of four components: student NSLP eligibility, parental education, books at home, and the percentage of NSLP-eligible students in the school the student is attending. We believe that using a sum score has value for reporting purposes as it allows stakeholders to clearly understand what the SES index points represent and what a specific expression on the index (say, 9 points) could mean in terms of the combination of certain items. Moreover, a principal component analysis showed similar empirical weights, and we found that while the four different components had somewhat different relationships with NAEP performance (see Appendix C), they were not sufficiently dissimilar to warrant assigning different weights on that basis. Therefore, we assigned the same point range (from 0 to 3) for the levels in each of the four components. The weighting scheme of the proposed NAEP SES index combines the features discussed in Cowan et al. (2012)—sum scoring that retains the meaning of the levels of the components, while the equal weighting chosen also has empirical support based on principal component analysis as well as the relationship of the components with the outcome variable (NAEP performance).

Importantly, the proposed four components have been collected by NAEP since 2003, and their collection is expected to remain reasonably stable for the time being. That is, there will be no need to frequently switch out one index component for another, as often happens for household possessions in PISA. While measurement invariance poses a significant challenge, the index nevertheless allows for tracking the relationship of NAEP performance to SES over the 20-year period from 2003 to 2022.

While issues of index construction and the weighting of the various components that go into it are certainly important, what has been missing, in our opinion, is a discussion of the criteria by which to judge the performance of an existing or new SES index, independent of its constituent items and the way it is scored and weighted. In the following section, we describe some of the guiding principles we used in developing the proposed NAEP SES index, as well as criteria we developed in partnership with NCES to evaluate its performance.

Guiding principles and criteria to evaluate the performance of the SES index

As exploratory work with different variants of a possible NAEP SES index has matured, in addition to research on different components and ways to combine them, we set out the following features as guiding principles:

First, the NAEP SES index should be useful for secondary research where it can be used as a reliable and valid SES control variable in education research to understand the unique effects of predictor variables of interest on educational outcomes.

Second, ideally, the SES index should be easy to understand for reporting purposes as well. In other words, the values of the SES index should be easily understood based on its constituent components. There may be different ways to obtain a certain index value based on the specific combination of the levels of the index components, but one should still have a sense of what the value represents.

Third, the SES index should work for all three grade levels—grades 4, 8, and 12—assessed by NAEP.

We also worked with NCES to define the following evaluation criteria for the proposed SES index (or alternative indices):

(1) How well does the NAEP SES index explain variance in NAEP performance?

Prior research has shown that SES plays a very important role in explaining variance in student academic outcomes (Sirin 2005). Similarly, the main function of the information collected by NAEP through questionnaires and other measures is to contextualize student performance and the differences in educational outcomes (NAGB 2013). We also believe that an SES index used for an educational assessment such as NAEP should capture aspects of SES that are most relevant to educational outcomes. Therefore, the NAEP SES index should explain a substantial amount of variance in student performance.

All else being equal, and as long as the index components are all SES-related, an index that explains more variance in student performance in NAEP would be one that captures more educationally relevant aspects of SES. At a minimum, the NAEP SES index should explain more variance in NAEP performance than is explained by NSLP alone (or any of the other SES-related components[7]). Also, the United States has participated in international large-scale assessments that employ different measures of SES while assessing similar student populations, thus providing us the opportunity to compare how well the NAEP SES index explains performance variation in NAEP compared to the extent to which other SES indices explain performance variation in their respective assessments. Finally, HSLS:09 allows us to compare the NAEP SES

[7] This is true for any single SES component in the proposed SES index, but we will only use NSLP as the comparison measure in the comparisons below because of its established status as an SES proxy reporting variable.

index with its own SES index, which is based on parent reports of the “big three” SES components.

(2) How well does the NAEP SES index account for achievement gaps[8] in NAEP?

If students from all racial/ethnic subgroups had on average very similar educational and economic opportunities, differences in average academic outcomes between them would not exist or would be small. However, the reality is that there are persistent differences in economic and educational opportunities between student subgroups, and differences in educational outcomes are therefore often quite large (Putnam 2016). Differences in SES should be a strong element in accounting for these persistent differences. Therefore, another basic criterion with which to evaluate the NAEP SES index will be if such an index, which is designed to capture a finer-grained picture of SES, can explain score gaps better than NSLP alone. As above, this criterion can also be compared to other SES indices employed in large-scale assessments administered in the United States with similar populations, grades, and subject levels.

(3) Does the index function similarly for major racial/ethnic subgroups?

It is well known that there are stark differences in SES between racial/ethnic groups in the United States.[9] Despite the different distribution of SES among student subgroups, the NAEP SES index should function, by and large, similarly across different racial/ethnic groups. That is, within each student subgroup, one should observe that higher SES index scores are associated with higher average NAEP scores, even though some student subgroups will have more or fewer cases in different score ranges of the NAEP SES index.

Overview of the Proposed NAEP SES Index

Components used in the proposed NAEP SES index and its scoring rules

Many years of research support the use of a NAEP SES index composed of several available components measured by NAEP. There have been several iterations and refinements over time, as the usefulness of certain potential index components were evaluated. For example, Broer, Xie, and Bohrnstedt (2017) and Xie and Broer (2017) originally proposed a five-component

[8] We use the term achievement gaps, score gaps, and educational outcome gaps interchangeably in this report. We see observed NAEP score differences between racial/ethnic subgroups as the product of a variety of existing inequities that are highly correlated with race/ethnicity, e.g., access to quality instruction and learning supports, opportunity to learn, availability of food and health resources, discrimination, among others. These factors are also related to SES. Therefore, a well-constructed SES index should explain at least some portion of the observed score gaps between racial/ethnic subgroups that are on average higher in SES and subgroups that are on average lower in SES.

[9] For a summary of literature on SES differences between racial/ethnic group in the US, one can refer, for example, to a summary by the American Psychological Association.

index with a subindex based on home possessions. Subsequent research by AIR and ETS placed doubt on the usefulness of the inclusion of the home possession subindex; thus, the currently proposed SES index includes only the four components described below.

It is beyond the scope of this summary report, which focuses on the currently proposed index, to describe all of the interim evaluations of possible additions or substitutions of certain components (e.g., the potential inclusion of the parental occupation variables that became available in 2017 or of technology items) and the specific findings. However, the common reasons for not including other items were (a) that they did not relate well to the four core SES components and/or (b) that they had a weak relationship to NAEP performance and thus their inclusion did not improve the index based on the evaluation criteria laid out above (and often made it perform less well). Moreover, for technology items, for which the above was true as well, another drawback is that the relationships these items have with other more stable SES items as well as with NAEP performance may change as time passes.

The following are the four components proposed for the NAEP SES index:

(1) **Number of books at home.** This variable comes from student responses to the question "About how many books are there in your home?" The response categories include *Few (0–10)*; *Enough to fill one shelf (11–25)*; *Enough to fill one bookcase (26–100)*; and *Enough to fill several bookcases (more than 100)*.

(2) **Student's eligibility for the National School Lunch Program (NSLP).** This variable represents student eligibility for NSLP based on school records. The original responses include *Eligible for NSLP*; *Not eligible for NSLP*; and *Information not available*. The category *Information not available* was treated as missing for this report. Thus, this variable is a binary variable.

(3) **Percentage of students eligible for NSLP at the school the student is attending.** This variable was originally derived from school responses to the question, "During this school year, about what percentage of students in your school was eligible to receive a free or reduced-price lunch through the National School Lunch Program?" Because of high missingness in the reported variable after a change in the school questionnaire in 2019, this variable is now primarily based on CCD data, but also includes school questionnaire responses when CCD data are missing (for a detailed explanation of the construction of the variable for this report, please refer to Appendix B). The categories are grouped into four levels: 0–25%, 26–50%, 51–75%, and 76–100%.

(4) **Highest level of education of either parent.** This variable is derived from students' responses to two background questions about the educational attainment of their mother and father and the highest level of educational attainment was chosen for either or both parents. The questions are "How far in school did your mother go?" and "How far in school did your father go?" The response categories include *Did not finish high school*; *Graduated from high school*; *Had some education after college*; *Graduated from*

college; and *I don't know*. The category *I don't know* was treated as missing in the SES index.

Each component is scored from 0 to 3, with lower values denoting lower SES and higher values denoting higher SES, creating an index that ranges from 0 to 12 points for grades 8 and 12 (Table 1). Student-level NSLP eligibility, which is a very important component of the index—an objective indicator of individual students' economic disadvantage—has only two levels, which are scored 0 and 3 to give this component an equal weighting vis-à-vis the other components in the index. Unlike in grades 8 and 12, parents' highest level of education is not collected in grade 4; therefore, the grade 4 SES index has only three components and ranges from 0 to 9.

Table 1. SES index components, categories, and scoring, by grade levels

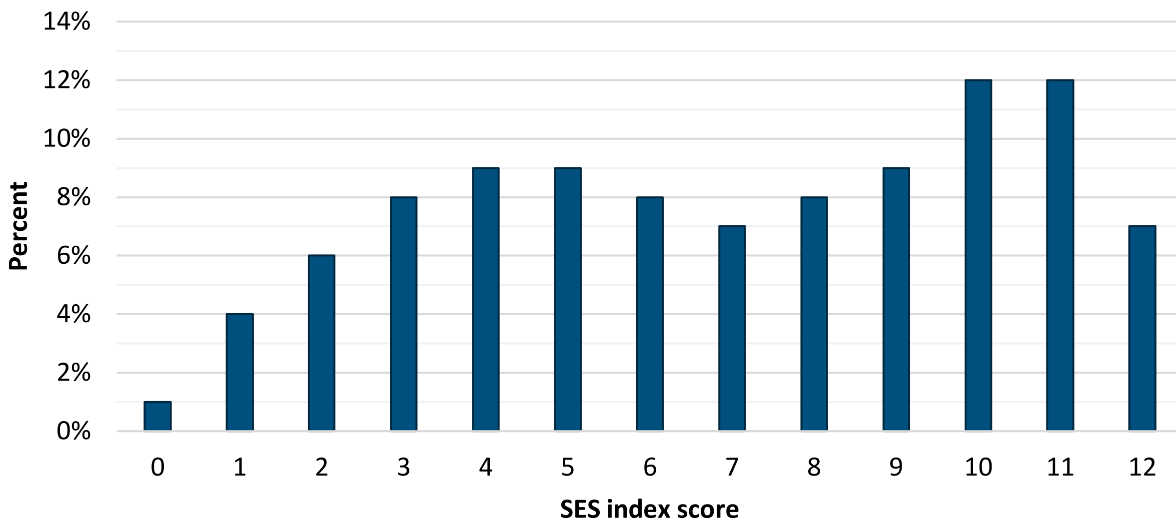
Components	Categories	Scoring	Grade levels
Number of books at home	0–10 books	0	Grades 4, 8, and 12
	11–25 books	1	
	26–100 books	2	
	More than 100 books	3	
Student's eligibility for NSLP	Eligible	0	Grades 4, 8, and 12
	Not eligible	3	
Percentage of students eligible for NSLP at school the student is attending	76–100%	0	Grades 4, 8, and 12
	51–75%	1	
	26–50%	2	
	0–25%	3	
Highest level of education of either parent	Did not finish high school	0	Grades 8 and 12
	Graduated from high school	1	
	Some education after high school	2	
	Graduated from college	3	

As SES is seen as a formative, multidimensional construct, the four components are therefore not expected to show very high correlations. However, they do show moderate intercorrelation (0.27–0.51) (see Appendix C, Table C.1), and the Cronbach's alpha coefficient denoting internal consistency was 0.69. This would be low for an index capturing a unidimensional construct based on indicator variables but is acceptable given the multidimensional nature of SES and the performance of the index according to the other criteria presented below.

Figure 3 shows the distribution of SES index scores for grade 8 mathematics for the full analytic sample. Note that the index scores are not normally distributed around the mean (this holds true for other grade levels as well). One could standardize the distribution, but this would remove the meaning of the index scores.

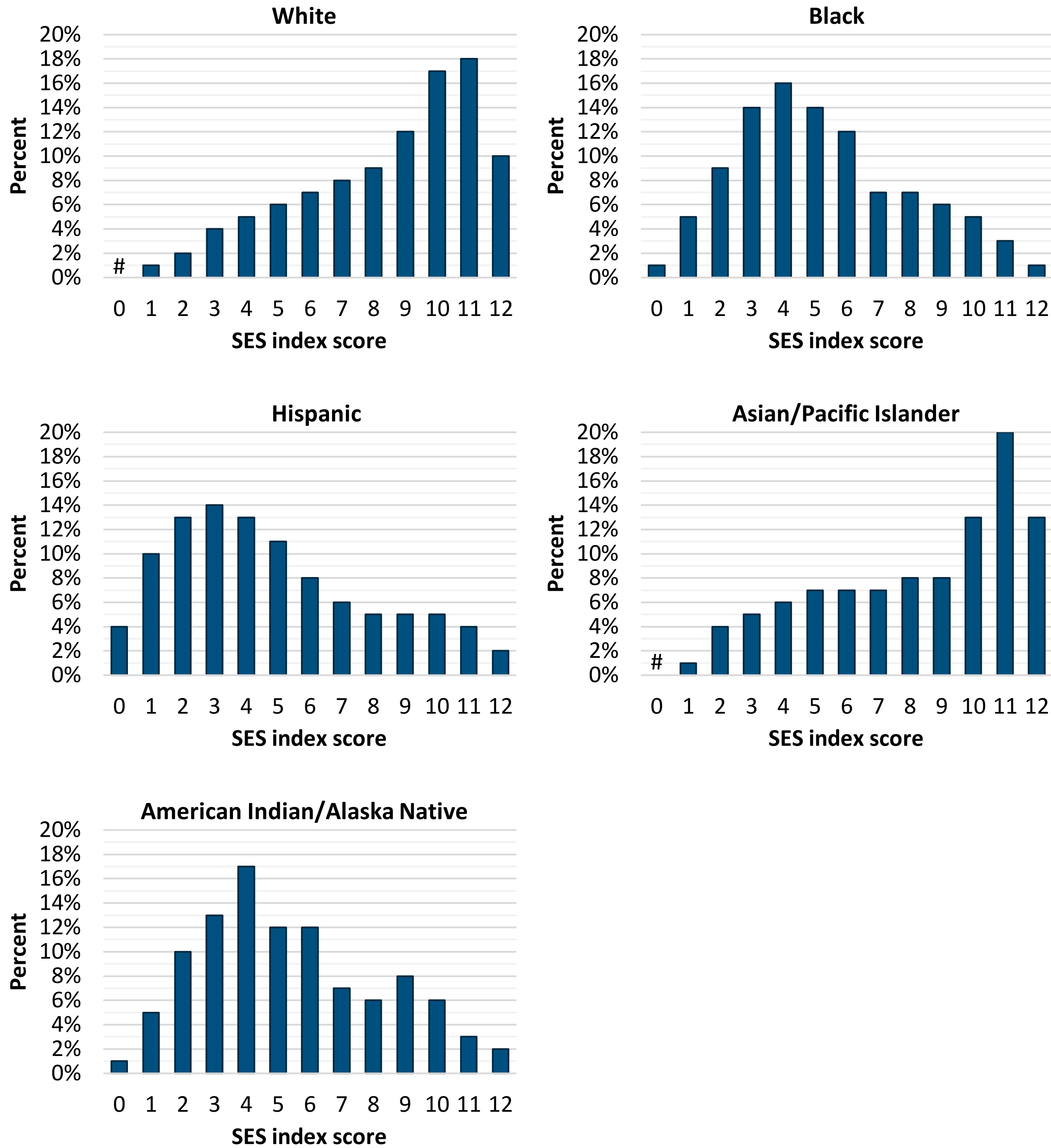
The reason for this bimodal distribution lies partially in the unequal SES distribution among racial/ethnic groups shown in Figure 4. White and Asian/Pacific Islander students have a relatively similar distribution, with more students with higher SES index scores (right-skewed distribution), while for Black, Hispanic, and American Indian/Alaska Native students the distribution is left-skewed—that is, more students have lower SES index scores. The NAEP SES index reflects the stark differences in opportunities that exist between different racial/ethnic groups in the United States.

Figure 3. Percentage distribution of SES index scores for grade 8 NAEP mathematics: 2019



NOTE: The SES index ranges from 0 to 12 points at grade 8. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Figure 4. Percentage distribution of SES index scores for grade 8 NAEP mathematics, by race/ethnicity: 2019



Rounds to zero.

NOTE: The SES index ranges from 0 to 12 points at grade 8. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. Students of Two or More Races assessed in grade 8 NAEP mathematics are not included in this figure due to small sample sizes. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

To evaluate the performance of the proposed index, we will review the results responding to the criteria established previously:

- How well does the NAEP SES index explain NAEP performance?
- How well does the NAEP SES index account for achievement gaps in NAEP?
- Does this index function similarly for major racial/ethnic subgroups?

Before reviewing these results, we will briefly review the analytic sample that was used to evaluate the NAEP SES index, using the example of 2019 NAEP grade 8 mathematics.

Description of the analytic sample

In this study, we used NAEP mathematics assessment data for grade 4, grade 8, and grade 12. Additionally, we conducted preliminary analysis for grade 8 reading and science assessment data and found similar results as were found for grade 8 mathematics, but the results of those analyses are not presented here. The plausible values of achievement scores and complex survey weights were taken into consideration in the analysis.

Using grade 8 mathematics assessment data as an example, a representative sample of grade 8 students in both public and private schools throughout the United States (and sometimes specified territories and possessions) was sampled for assessment through a multistage sampling design. For our analyses, we obtained the analytic sample for each subject and grade level by restricting the analysis first to the national reporting sample and then to the 50 states and the District of Columbia, to the national public school sample, to the states with valid questionnaire information for the key analytic variables, and, finally, to the cases with no missing values on the key analytic variables. Please refer to Appendix A for more details about how the complete case analytic sample[10] was obtained through each step.

How well does the NAEP SES index explain NAEP performance?

One way to assess how well the SES index functions in explaining NAEP performance, which is more intuitive than the R^2 figures presented below, is to examine how each additional SES index score is associated with higher average NAEP performance.

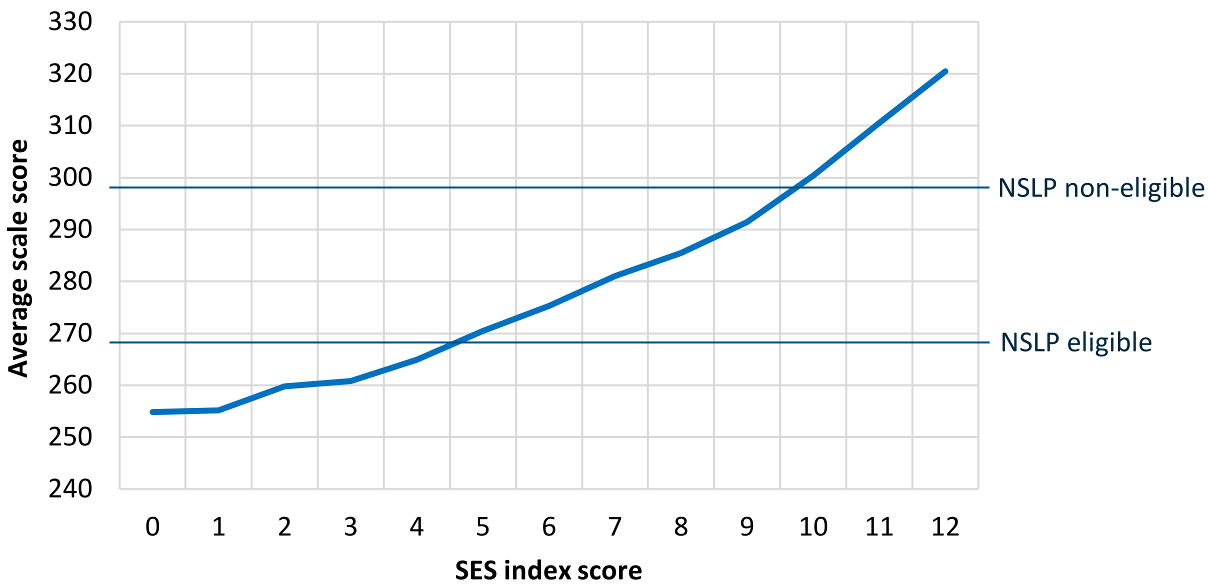
As shown in Figure 5, one can see that each additional SES index point is associated with higher average NAEP math performance (although the increases between index points 0 and 1 and

[10] Previous research with the SES index has used fully imputed datasets (e.g., Bai, Straus, and Broer 2021). We decided to use a complete case sample approach here as it makes the analysis and presentation of results more straightforward and makes it easier for researchers to replicate or build upon the findings in this report. (The challenge of dealing with missing data is covered later in the report.) Missing data are not at random but are likely connected to SES levels and performance. It is therefore possible that many of the excluded students are from lower SES backgrounds. Had they been included through imputation of their missing values, the results presented below (e.g., on explaining NAEP performance or accounting for achievement gaps) could have been even stronger. We therefore view the results presented below as conservative.

between 2 and 3 are small). Please also refer to Figure F.3 in Appendix F for a box plot of grade 8 mathematics scale scores for each SES index score.

By showing the locations of average scores for NSLP-eligible and non-eligible students in Figure 5, one can also see the value of having a finer-grained measure of SES, with index scores that are associated with a range of average NAEP scale scores, including some that fall below the average scale score for NSLP-eligible students (268), some that fall above the average scale score for NSLP-non-eligible students (298), and some that fall in between. The SES index scores, and their associated average NAEP scale scores, range from below the *NAEP Basic* performance level at the low end of the SES index (0 to 2 points) to midway between *NAEP Proficient* and *NAEP Advanced* for the highest NAEP SES index scores (11 to 12 points).[11]

Figure 5. Average scale score for grade 8 NAEP mathematics, by SES index score and NSLP eligibility: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. The SES index ranges from 0 to 12 points at grade 8. The information about the National School Lunch Program (NSLP) variable is based on available school records. If school records were not available, the student was classified as "Information not available." Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

While this graph is valuable in understanding how the NAEP SES index functions, we will now present numerical evidence that can be used to compare the relationship between the SES index and NAEP performance vis-à-vis the relationship between NSLP alone and NAEP

[11] The achievement-level cut points for grade 8 NAEP mathematics are 262 points for *NAEP Basic*, 299 points for *NAEP Proficient*, and 333 for *NAEP Advanced*. For descriptions and more information about NAEP achievement levels, refer to https://nces.ed.gov/nationsreportcard/mathematics/achieve.aspx.

performance, as well as to compare this relationship in NAEP to those in similar large-scale assessments.

First, we conducted binary regressions to examine how well the SES index explains students' performance compared to using NSLP status alone. Table 2 shows that the SES index performs better than NSLP alone in terms of explaining the variance in grade 4, 8, and 12 students' performance in mathematics.[12] For example, for grade 8 students, the variance in mathematics performance (R^2) explained by the SES index is 0.24 compared to 0.14 explained by NSLP alone, suggesting that the SES index explains an extra 10 absolute percentage points of variance (relatively 70% more) in students' mathematics achievement. These results are consistent with the results obtained for grade 4 and grade 12 mathematics. The SES index explains an extra 7 percentage points of variance in fourth-graders' achievement and an extra 11 percentage points of variance in twelfth-graders' achievement. (For additional information, please refer to Table D.1 in Appendix D.)

Table 2. Variance explained by NSLP and SES index for grade 4, 8, and 12 NAEP mathematics performance: 2019

	Grade 4 Mathematics		Grade 8 Mathematics		Grade 12 Mathematics	
NAEP 2019	NSLP	SES	NSLP	SES	NSLP	SES
R^2	0.14	0.21	0.14	0.24	0.11	0.22

NOTE: The R^2 values in this table are calculated based on binary regressions using the SES index as a continuous variable. The NAEP mathematics scale ranges from 0 to 500 at grades 4 and 8 and from 0 to 300 at grade 12. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." Standard errors are shown in parentheses. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment.

Comparison to SES indices in other large-scale assessments

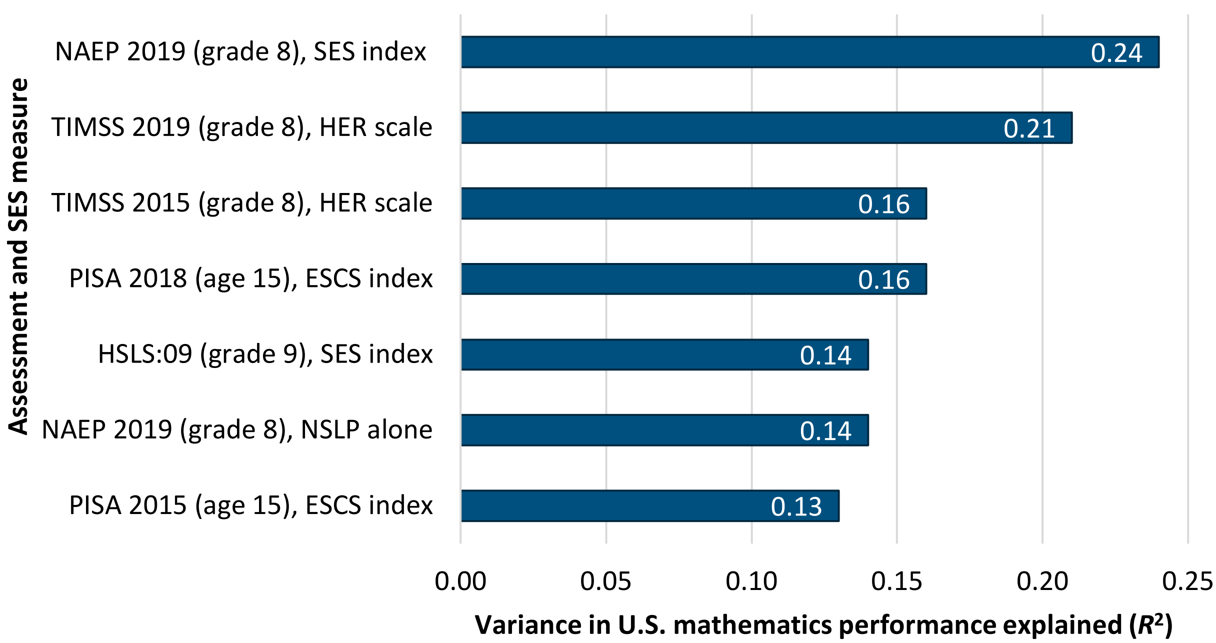
U.S. students are assessed in the same general subject areas in different but related assessments. For example, the Trends in International Mathematics and Science Study (TIMSS), like NAEP, assesses U.S. students in grade 4 and 8 mathematics. Of course, the sampling framework, assessment framework, and assessments of TIMSS differ from NAEP's, but it can still be useful to compare the variance explained by the TIMSS SES-related index (the home educational resources, or HER, index) in grade 8 mathematics with the variance explained by the proposed NAEP SES index for grade 8 mathematics. We have shown above that the proposed NAEP SES index explains performance better than NSLP alone, but how does it perform compared with similar SES indices used by other assessments?

[12] Preliminary analyses not shown here suggested similar findings for grade 8 science and reading, but with a lower R^2 value for reading than those for mathematics and science.

Figure 6 presents the results for TIMSS grade 8 mathematics (both for 2015 and 2019) and for the PISA mathematics assessment (both for 2015 and 2018), which assessed 15-year-olds and used an index of economic, social, and cultural status (ESCS). Moreover, the figure includes a comparison with HSLS:09, which assessed grade 9 students with an algebra-focused assessment and used a parent-reported SES index of the "big three" SES components (parental education, occupational prestige, and income). The variance in grade 8 NAEP mathematics explained by NSLP alone is displayed for comparison, as well.

We see that the NAEP SES index performs well compared to the SES-related indices used in other assessments in terms of variance explained in mathematics-related performance for similar age groups. Interestingly, the HSLS:09 SES index, based on parent reports of the big three SES components, did not perform better in this metric than NAEP NSLP alone, a finding that presents further evidence about the continued importance of including NSLP as an income proxy as part of an SES index.

Figure 6. Variance in U.S. mathematics performance explained, by assessment and SES measure



NOTE:

1. The R^2 values in this figure are calculated based on binary regressions using the NAEP SES index as a continuous variable. The information in NAEP about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available."
2. The home educational resources (HER) scale in TIMSS includes number of books in the home; number of home study supports (Internet connection or own room or both); and highest level of education of either parent, which is derived from both the father's and mother's highest educational levels.
3. The economic, social, and cultural status (ESCS) index in PISA is a composite score of parents' highest level of education, parents' highest occupational status, and home possessions.
4. The SES composite variable in HSLS:09 is calculated using parent/guardian reports of their education, occupation, and family income.

For more detail about the SES-related indices in TIMSS, PISA, and HSLS:09, refer to Appendix G.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment; High School Longitudinal Study of 2009 (HSLS:09); Program for International Student Assessment (PISA) 2015 and 2018 Mathematics Assessment; and Trends in International Mathematics and Science Study (TIMSS) 2015 and 2019 Grade 8 Mathematics Assessment. See also Yin and Fishbein (2020) for TIMSS; Organization for Economic Cooperation and Development (OECD) (2019) for PISA; and Ingels et al. (2014) for HSLS:09.

How well does the NAEP SES index account for achievement gaps in NAEP?

One of NAEP's key roles as the Nation's Report Card is to report and track educational outcomes for major racial/ethnic groups. As mentioned earlier, differences in SES should be a key factor in explaining these persistent differences in outcomes between racial/ethnic subgroups in the United States. Therefore, we examined the achievement gaps (or educational outcome gaps) between different racial/ethnic groups, before and after controlling for limited English proficiency (LEP) and either NSLP or the NAEP SES index.

Four analytic models were run:

- Model 1 was the null, or baseline, model, showing the achievement gaps between subgroups without controlling for any variables.
- Model 2 added LEP as a control variable, as it is an important variable for better understanding outcome differences between White and Hispanic students.
- Model 3 further added students' eligibility for NSLP to Model 2 as an SES control.
- Model 4, as a comparison model, added students' SES index scores (instead of NSLP) to Model 2.

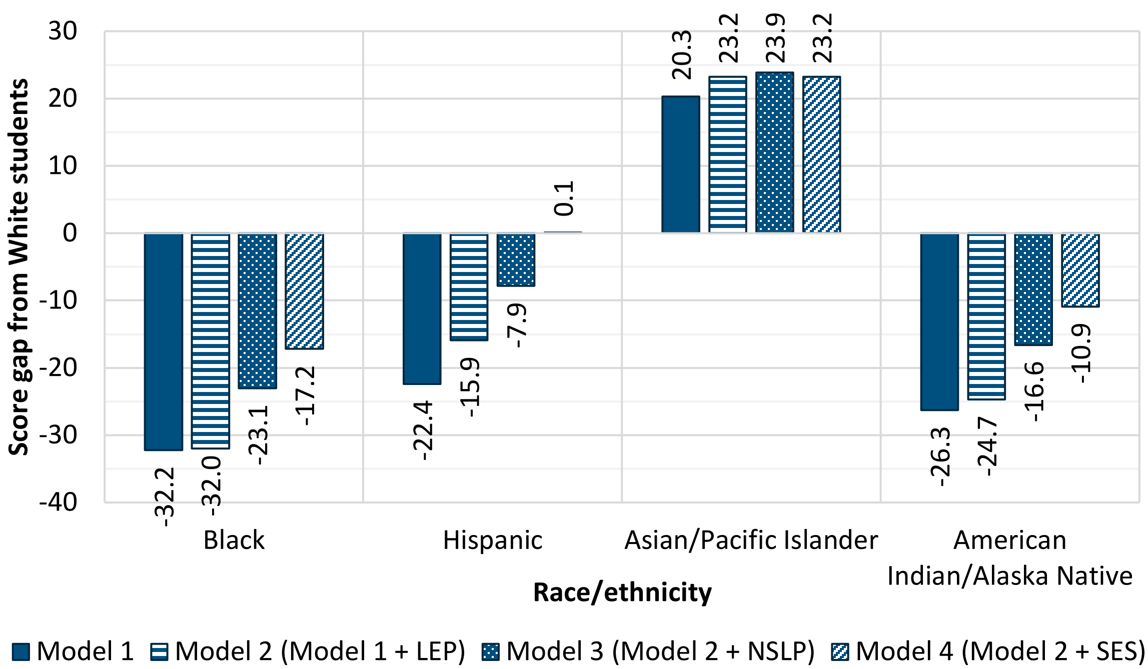
Figure 7 and Figure 8 present the results of these models for the grade 8 and grade 4 NAEP mathematics assessment data, respectively (see also Appendix E, Table E.1 and Table E.2). Given the findings are similar between grade 4 and grade 8, we only discuss the results for grade 8 with model comparisons.

Figure 7 shows the educational outcome gaps between White students and other racial groups in each model for the grade 8 mathematics assessment data. For example, the gap between White and Black students is 32.2 NAEP points in Model 1, indicating that the average score of Black students is 32.2 points lower than that of White students. Model 2 indicates that English proficiency is not a key factor in explaining this gap, since the gap remained as 32.0 scale points after controlling for LEP status. Model 3 added students' NSLP status and the remaining unexplained gap was 23.1 scale points, indicating that differences in NSLP status between Black and White students account for some of the score gap. Lastly, Model 4 added the SES index into the model (instead of NSLP), and the remaining unexplained achievement gap was 17.2 scale points, which is almost 47% lower than the observed gap in Model 1. This indicates that the SES index explains a larger amount of the Black-White achievement gaps in NAEP than is explained by NSLP alone. This pattern holds true for the American Indian/Alaska Native-White achievement gap, as well. (The results are presented with details in Appendix E.)

The results for the Hispanic-White achievement gap tell a different story. In Model 2, English proficiency shows its importance in explaining the score difference—the gap was 15.9 scale points once LEP was included, representing a 29% reduction from the Model 1 gap of 22.4 scale points. The comparison between Model 3 (−7.9 points) and Model 4 (0.1 points) indicates that the SES index has additional power in explaining the Hispanic-White achievement gap compared to NSLP alone. Another way to look at it is that the Hispanic-White score gap is fully explained by differences in students' English proficiency and SES background.

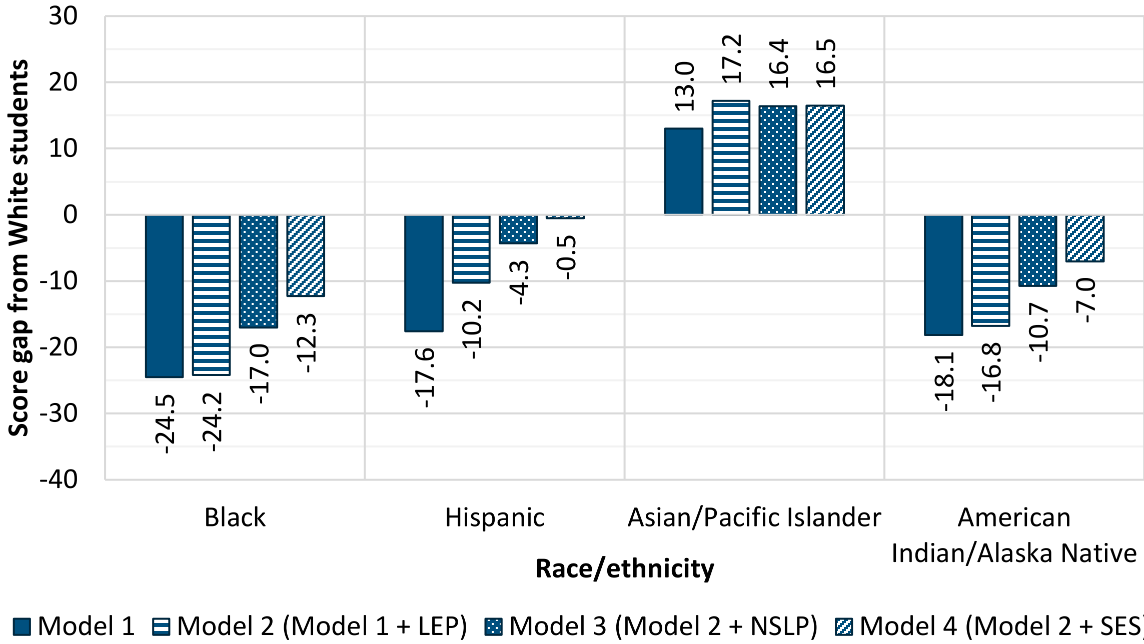
By contrast, the Asian/Pacific Islander-White achievement gap favors Asian/Pacific Islander students by 20.3 NAEP points (Model 1), and including language proficiency, NSLP, or the NAEP SES index does not help explain the gap. Rather, the gap becomes larger in Models 2–4 by approximately 3 NAEP points, suggesting that Asian/Pacific Islander students' NAEP average would be 3 NAEP points higher than White students' (i.e., 23.2 points higher rather than 20.3 points) if differences in English proficiency and family background were held constant. Future work should focus on investigating potential factors or mechanisms behind this large score gap, which remains unexplained by SES.

Figure 7. Score gaps between White students and other racial/ethnic groups for grade 8 NAEP mathematics, by model: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. Students of Two or More Races assessed in grade 8 NAEP mathematics are not included in this figure due to small sample sizes. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Figure 8. Score gaps between White students and other racial/ethnic groups for grade 4 NAEP mathematics, by model: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 4. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. Students of Two or More Races assessed in grade 4 NAEP mathematics are not included in this figure due to small sample sizes. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." Some apparent differences between estimates may not be statistically significant.

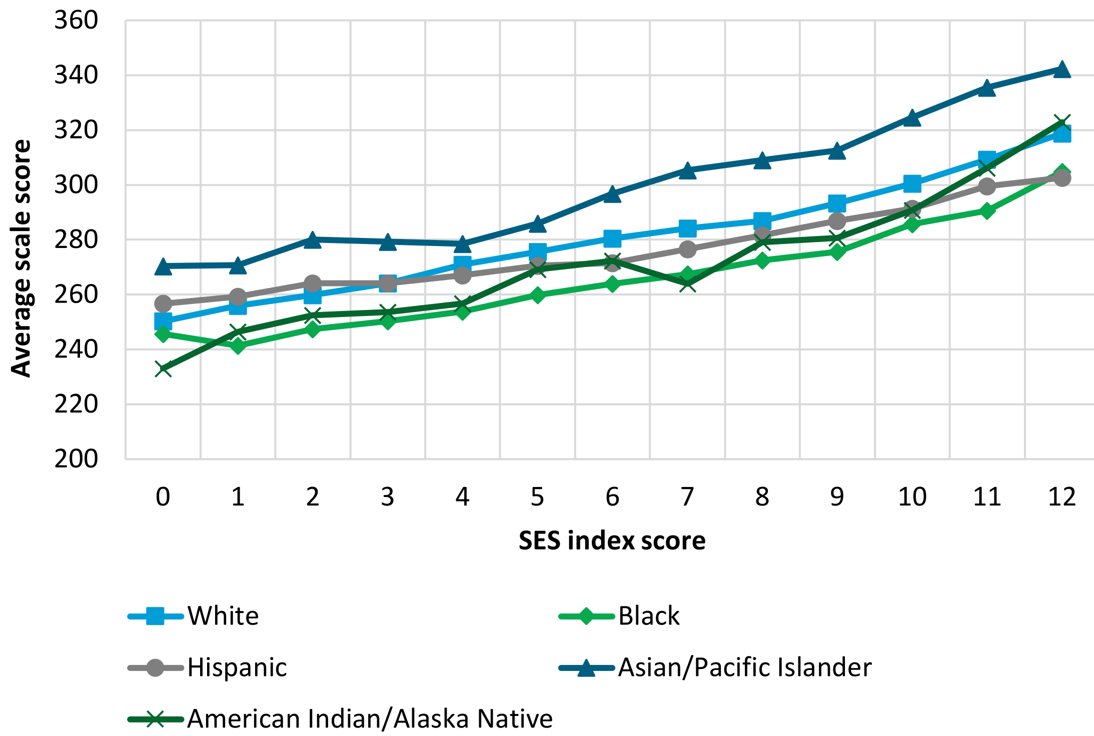
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4 Mathematics Assessment.

Does the NAEP SES index function similarly for major racial/ethnic subgroups?

While there are strong SES differences between racial/ethnic student subgroups, a NAEP SES index should nevertheless function in similar ways for the different student subgroups, at least in the sense that higher SES index scores would be associated with higher average NAEP scores within each racial/ethnic subgroup, by and large.

Therefore, we show below the average achievement score by each SES index score for each racial/ethnic group. Figure 9 shows that the SES index functions similarly for each subgroup using the grade 8 mathematics assessment as an example. The relationship between SES index score and average NAEP performance looks somewhat different for American Indian/Alaska Native students; while we do see a generally rising trend it is not as smooth as the other lines and there are some dips within the generally rising function between SES index points and average NAEP scores which is probably due to the lower sample sizes for this subgroup (see Appendix F for the corresponding figures for grade 4 and grade 12).

Figure 9. Average scale score for grade 8 NAEP mathematics, by SES index score and race/ethnicity: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. Students of Two or More Races assessed in grade 8 NAEP mathematics are not included in this figure due to small sample sizes. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Please note, that we do not claim here that there are no interaction effects between race/ethnicity and the SES measure at all (which would be an overly strict criterion), just that the SES index seems to be functioning, by and large, in a similar manner across major racial/ethnic subgroups.

In summary, the proposed NAEP SES index performs well with respect to all three established evaluation criteria.

Examples of How the NAEP SES Index Has Been Used

The following are two recent examples of secondary analyses where the proposed NAEP SES index, or a variation of it, has been used.

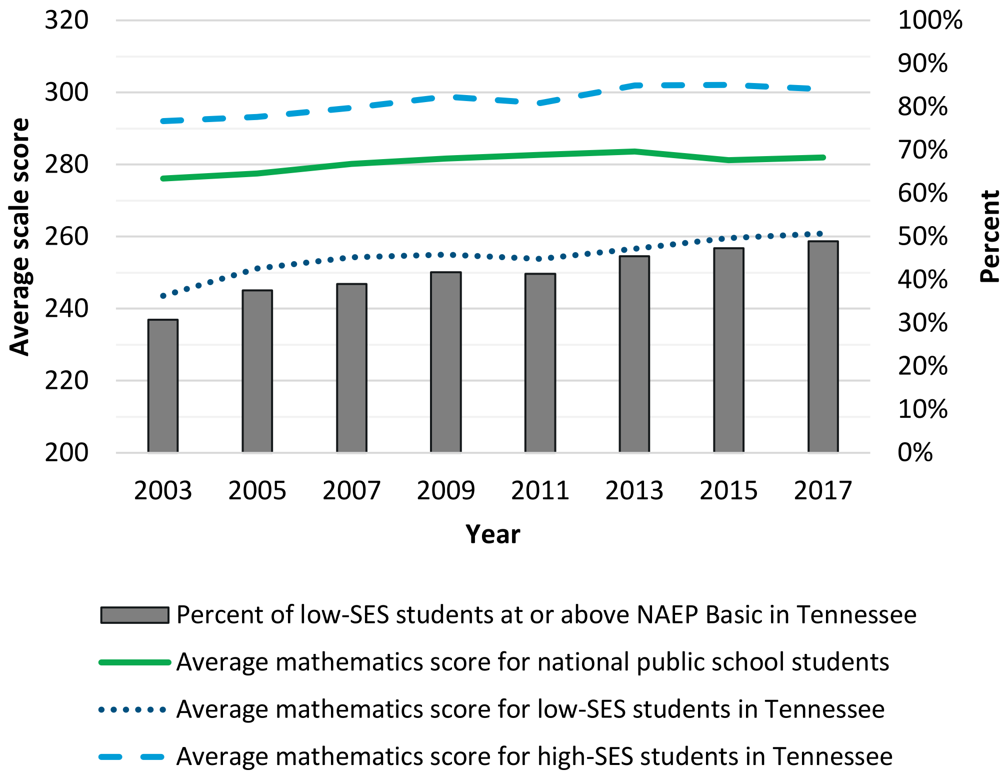
U.S. national and state trends in educational inequality due to socioeconomic status: Evidence from the 2003–17 NAEP

The SES index presented here can be used for NAEP trend reporting or analysis starting from 2003. There has been a growing interest in understanding trends in SES achievement gap in the United States and other countries using large-scale educational assessments (Chmielewski 2019; Hanushek et al. 2022; Reardon 2011, 2021; Broer, Bai, and Fonseca 2019). In a recent working paper, AIR researchers used the SES index proposed here to further examine national and state trends in educational inequality due to family SES, asking:

- Has the socioeconomic achievement gap changed over time nationally in the United States and by state?
- Has the performance of low-SES students improved over time?

Using the proposed SES index, students were grouped into top and bottom quartiles of SES in each state each cycle of the NAEP grade 8 mathematics assessment between 2003 and 2017. The achievement gap between students in the top and bottom quartiles of SES was then calculated for each state in each administration year. The results showed that the national SES achievement gap remained unchanged over this time period and that SES achievement gaps showed no significant change in 34 states while widening in 14 states and narrowing in only two states. Tennessee is one of the states that showed a narrowing SES achievement gap, with both high- and low-SES students improving their academic performance while low-SES students improved at a faster rate (Figure 10).

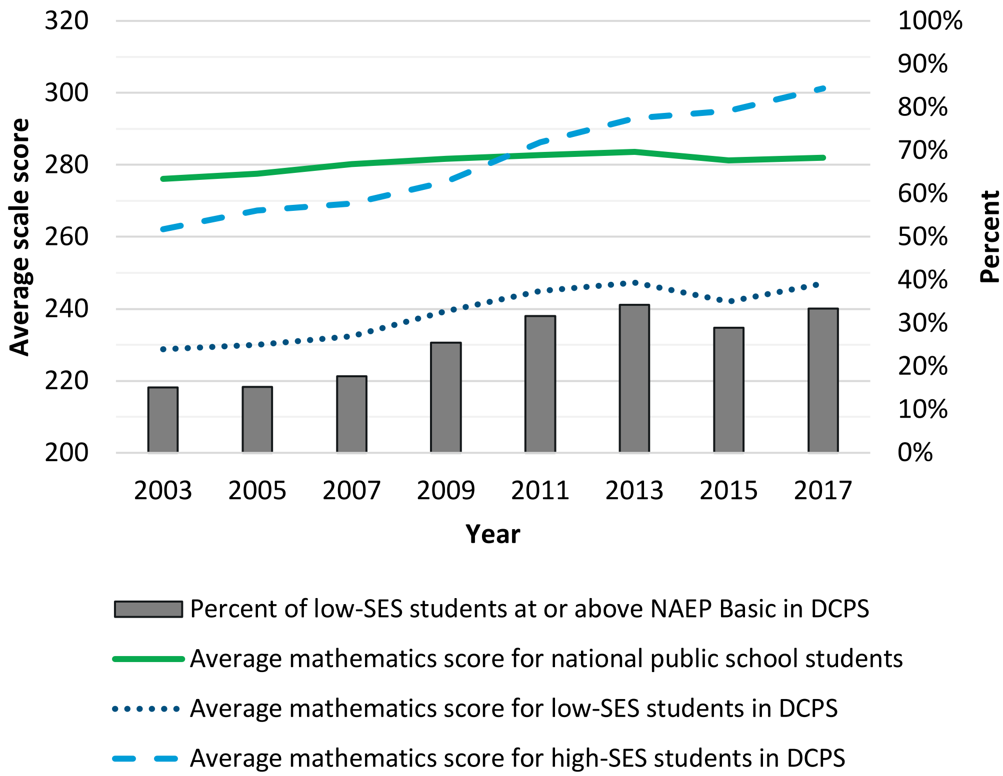
Figure 10. Percentage of low-SES students performing at or above *NAEP Basic* and average score in grade 8 NAEP mathematics by SES background, Tennessee: 2003–2017



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. The SES index ranges from 0 to 12 points at grade 8. Low-SES students are defined as students in the bottom 25% of the SES distribution and high-SES students are defined as those in the top 25%. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, 2007, 2009, 2011, 2013, 2015, and 2017 Grade 8 Mathematics Assessments.

A contrasting example would be the trend seen in the District of Columbia Public Schools (DCPS). Figure 11 shows that the average performance of all students in DCPS has continuously risen since 2003, getting closer to the national public school average. However, the strong improvement in average achievement is only part of the story. DCPS's gap between high- and low-SES students in the same time period increased from 33 points in 2003 to a substantially higher 54 points in 2017. The SES achievement gap started widening especially rapidly after 2009, as high-SES students improved their performance much faster than low-SES students.

Figure 11. Percentage of low-SES students performing at or above *NAEP Basic* and average score in grade 8 NAEP mathematics by SES background, District of Columbia Public Schools (DCPS): 2003–2017



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. The SES index ranges from 0 to 12 points at grade 8. Low-SES students are defined as students in the bottom 25% of the SES distribution and high-SES students are defined as those in the top 25%. Differences were calculated using unrounded values. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, 2007, 2009, 2011, 2013, 2015, and 2017 Grade 8 Mathematics Assessments.

In summary, the proposed SES index allows for an investigation of changes in educational inequality due to family SES in the United States over a long time period. As such, the proposed SES index used in the study contributed to the literature not only by reflecting on national trends, but also by providing state-level examinations.

Tracking the progress that states have made toward more equitable educational outcomes between lower and higher SES students can also help identify those jurisdictions that have had more success in reducing SES gaps, allowing others to learn from their approaches and spurring additional efforts in places where SES gaps are widening instead of narrowing.

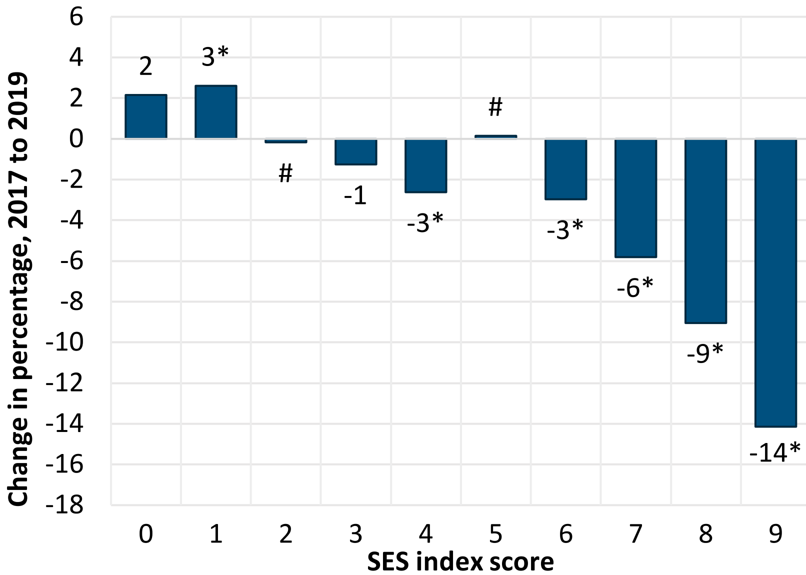
Trends in grade 4 smartphone access by SES index score: 2017 to 2019

A variant of the SES index discussed in this report has also been used to examine changes between 2017 and 2019 in students' reported access to technology items at home (Shipan and Broer 2022). This analysis was conducted with the goal of understanding technology trends among students, particularly given the growing importance of technology in education as well as because technology items at home have been considered as possible SES proxies in the past.

One of the most striking findings from this investigation was that changes in smartphone access varied at grade 4, particularly by SES index score. As presented in Figure 12, some low-SES students reported greater average access to smartphones in 2019 than in 2017, while high-SES students saw sharp decreases in access across the same time period. For example, the percentage of grade 4 students reporting smartphone access at home increased by 3 percentage points for those with an SES index score of 1 point but decreased by 14 percentage points for those with the highest SES index score (9 points). This analysis provides an example of how index-point-by-index-point comparisons may be helpful for diagnostic purposes from one NAEP administration to the next.

The analysis is also relevant for the question of incorporating technology items as components in an SES index. While other SES-related indices (e.g., PISA's ESCS index) include home possessions such as digital devices (OECD 2019), these results serve as a caution against using technology items at home as a proxy for SES in NAEP. The findings indicate that the relationship between technology access and socioeconomic background may not be as straightforward as some have assumed and that it may be prone to strong change even over short time periods.

Figure 12. Change in percentage of public school students in grade 4 NAEP mathematics who report having access to a smartphone at home, by SES index score: 2017 to 2019



Rounds to zero.
* $p < .05$. Change in percentage from 2017 to 2019 is significant at the .05 level of statistical significance.
NOTE: The data in this figure represent the change from 2017 to 2019 in the percentage of students in each SES index score group who reported having a smartphone that they can use at home. Parents' highest level of education is not collected in grade 4; therefore, the grade 4 SES index has only three components and ranges from 0 to 9. The reported school-level NSLP variables (C051651 in 2019 and C051601 in 2017) had high missing rates. Thus, for schools with missing data, the percentage of students eligible for NSLP at the school was calculated from the student-level NSLP variable (SLUNCH1) as follows: number of NSLP-eligible students at the school divided by the total number of sampled students at the school. Schools with 25% or more students coded as SLUNCH1 = 3 "Info not available" were excluded from the analysis. Schools with no NSLP-eligible students and one or more students coded as SLUNCH1 = 3 "Info not available" were also excluded from the analysis. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2017 and 2019 Grade 4 Mathematics Assessments.

Description of Challenges

While we believe the evidence presented so far confirms that the proposed NAEP SES index functions well according to the criteria established and that it would have utility both for reporting and for research applications, we understand that challenges remain, especially with respect to its operational implementation in NAEP. Some are complex issues that we will not be able to cover here in detail, but we hope to provide a high-level summary of the main challenges.

Comparisons across time

By choosing SES index components that have been relatively stable across time (the four components have been collected since 2003), the proposed index avoids one issue that other comparable indices have to deal with: that is, how frequently to retire components and introduce new ones over time. This is a relevant issue in PISA, which uses many household items in its ESCS index and changes them relatively frequently.

However, even if index components stay the same, their meaning or relative importance can change over time. Each stand-alone item can have arbitrary categories and suffer from potential changes in the meaning of its categories, both over time and/or across jurisdictions. For example, in Figure 2, the SES proxy variable “highest level of parental education” was used to examine the trend in mathematics score gaps between students whose parents “graduated from college” and those whose parents “graduated from high school” over time. While there is often not much change over short periods, these categories can convey different meanings over longer periods.

Thus, in the comparison between 1990 and 2019, with the expansion of higher education in the intervening time, more families now have one parent whose highest education is a college degree. That means more students will fall in the highest parental education category in 2019 than they did in 1990. The same changes can also be seen in other variables over longer time periods (e.g., fewer students selecting the highest category in the books at home variable,[13] a rise in the percentage of NSLP-eligible students). These changes over longer time periods (as well as differences in overall SES levels across jurisdictions) can be seen as an argument against defining absolute cut-off points on the NAEP SES index (e.g., denoting a “high” SES level as 10 points and above; “middle” as between 5 and 9 points; and “low” as 4 points and below). While on the one hand, such definitions may add additional context to reporting, on the other hand, establishing—and revisiting SES cut points over time—may prove to be contentious.

Concerns about the robustness of the SES index over time merit continuous validation efforts. One way to address concerns is to be open about the fact that small continuous changes are expected to happen and that over longer time periods, these changes can become larger, making comparisons based on certain SES index scores, or score categories tied to specific cut-off points, less advisable. This could mean implementing different reporting methods based on the time horizon for the comparisons: concurrent reporting, administration-to-administration reporting, and trend reporting.

For **concurrent reporting**, there is no concern about changes in the meaning of the SES components: for a given year, say, NAEP 2024, one can use the SES index to provide a fine-grained contextualization (index point by index point) of the different SES circumstances of different student groups (and in different states) and how they relate to average NAEP performance. In addition, one could choose to report by previously defined categories based on cut-off points (e.g., high, middle, low).

For **administration-to-administration reporting**, changes in the underlying SES variables should be minor (especially for the grade 4 and 8 mathematics and reading administrations, which

[13] However, the relationship between books at home and NAEP performance has stayed fairly stable over time.

occur every 2 years, or every 3 years in the case of the 2019 and 2022 administrations, due to COVID-related delays). This means that after the appropriate validation checks, it should be possible to use index-point-by-index-point reporting for specific reporting purposes: for example, to examine score changes from 2019 to 2022 by NAEP SES index point (which, with sufficient sample size, could be broken down further, for example, by major racial/ethnic subgroup). Such reporting could potentially enhance our understanding of the circumstances associated with smaller or greater score declines (or gains) in the future. This could also be helpful for diagnostic purposes, either for internal checks or for specific reporting purposes. (See the above example on the change in fourth-graders' reported smartphone access by SES index point from 2017 to 2019.)

For **trend reporting**, reporting over longer time periods by each individual SES index score would be more challenging because small changes that may not be problematic over a shorter time span can accumulate over time. As a consequence, while the NAEP SES index can be produced back to 2003 for trend reporting, we would not recommend interpreting changes between 2003 and 2019 index-point-by-index-point or by absolutely defined SES categories. Rather, we would propose using relative categories instead of absolute ones, as we did in the research study on trends in SES achievement gaps at the national and state levels mentioned in the introduction (Bai, Straus, and Broer 2021). In that study, we did not define absolute levels of SES for reporting categories (e.g., 10 points and above is high SES) and use them to track performance gaps between low- and high-SES students over time. Instead, we proposed dealing with potential changes in SES over time by looking at high-SES students defined by a certain percentage in the SES distribution (e.g., top 25%) and comparing their performance to that of lower-SES students (e.g., bottom 25%).

Independent of the changes in the NAEP SES index distribution, one can always identify students at the upper and lower ends of the distribution for each year.[14] This approach is also helpful for state-level trend reporting. While analyzing average NAEP scores by SES points or established SES levels (defining absolute points on the SES scale as low, middle, and high SES) can show differences in SES distributions across states, which can provide useful contextual information in concurrent or administration-to-administration reporting. Another approach could be to use state-specific metrics, defining relatively lower- (e.g., bottom 25%) and higher- (e.g., top 25%) SES groups for each state, and tracking the difference in performance in each state between students in these two groups over time (Bai, Straus, and Broer 2021).

Another example of relative reporting is the comparison of performance changes by percentile groups, which has enhanced our understanding about the evolution over time of achievement gaps between subgroups based on different percentile ranks of the score distribution. This comparison has been used in reporting on score changes between two administrations (as portrayed in the Nation's Report Card) and also over longer time periods (Burg et al. 2022).

[14] For a method to always compare groups of exactly 25%, please refer to Bai, Straus, and Broer (2021).

The NAEP index proposed here employs a sum score approach, which makes it easier for audiences to understand how it is formed and what the different values represent. We see this as a plus. While it is unlikely that more complex and less easy to understand forms of index construction (standardization; weights based on principal component analysis) would improve the performance of the SES index, it is nevertheless possible that other considerations need to be made for operational implementation. One such consideration could be the ease of changing index components in the future, which may require either a different approach to the sum score approach described here or a hybrid approach where the sum score approach could be used in current reporting and administration-to-administration reporting, but a different approach could be used for reporting longer-term trends. In any case, we suggest that longer-term trend reporting might be better conducted on a relative basis (comparing score gaps between students at the 25th percentile vs. the 75th percentile of SES index distribution).

Missing data treatment

Another challenge concerns how to deal with missing data in the SES index's various components. In this report, we use a complete case sample, which allows researchers to more easily replicate the results reported here. In other investigations regarding the validation of variants of the NAEP SES index (Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017) or its application in a research study (Bai, Straus, and Broer 2021), we used multiple imputation to impute values for missing components. A recent research study also looked at issues regarding imputation of the school NSLP percentage (based on school questionnaire data) and found that multilevel multiple imputation implemented with the *mice* R package (van Buuren and Groothuis-Oudshoorn 2011) was able to approximate "true values" even when 50% missingness was simulated using a complete case data set (Yavuz et al. 2022).

The missing rates for books at home, school NSLP percentage (with the methodology described in Appendix B, i.e., based primarily on the CCD), and individual student NSLP eligibility (at least for public school students) are reasonable, and imputation would not be expected to generate concerning imputation errors. However, missing data on parental education is more problematic. As Jewsbury et al. (2018) noted, there are several different missingness types to consider: normal item-level missingness within the questionnaire (e.g., skipping a question), states opting out of the student questionnaire, and students who answer "I don't know" to the question. The mechanisms for dealing with these three types of missingness are different and may require different treatment. In general, students answering "I don't know" tend to have low NAEP scores that are similar to those of students who report that their parent(s) did not finish high school, meaning the missingness is not at random.[15]

Another methodological and practical issue that needs to be resolved is what type of imputation methodology to employ. Multiple imputation is well equipped to deal with the

[15] Jewsbury et al. (2018) also noted that missingness for Hispanic students in that variable is three times higher than for other racial/ethnic subgroups.

problem, given its favorable performance and its handling of uncertainty in the imputation process (Rubin 1987) by providing not one, but several imputed values. This methodology is also well known for its flexibility in handling different types of variables (Hughes et al. 2014). But given that multiple imputation produces several values per student for a missing value that have to be taken into account jointly in analysis, this can also be seen as a practical limitation. Up to now, NAEP has never used multiple imputation on the contextual questionnaire data. Thus, employing a multiple imputation approach would require additional documentation for data users.

One could reduce the different imputed values for each variable to a single imputed value for each student (average or mode of several imputed values). Because such an approach would remove information about uncertainty in the imputation process, it has clear drawbacks. However, in early research on a variant of the proposed index, we applied weighted sequential hot deck imputation and computed the rounded average of 20 SES index values for subsequent analyses. We found that the main result (the relationship with NAEP performance, accounting for the achievement gap) was similar for both a complete case model and the fully imputed dataset (Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017).

Additional research and discussion about the pros and cons of different methods most appropriate for operational implementation are needed and will continue to be explored.

The NAEP SES index in grade 4

Another challenge is that the grade 4 SES index is comprised of only three components because parental education is not collected from fourth graders (see Table 1). Even though this means that a major SES component is not represented in the index at grade 4, previous research (Jewsbury et al. 2018; Ogut, Bohrnstedt, and Broer 2016) has shown that books at home is more closely related to parental education than it is to the income component of SES (and could therefore act as a proxy for the missing component). While the resulting internal consistency measures are lower as a consequence of having one fewer component in the index, the overall amount of variation in NAEP scores explained by the three-component SES index in grade 4 (R^2 = 0.21) is not very different from those associated with the four-component index used in grades 8 and 12 (R^2 = 0.24 and 0.22, respectively).

If it were deemed desirable to have a four-component index with a different fourth component in grade 4, one could consider including a subindex of technology items in the home. Jewsbury et al. (2018) found that these technology items form a distinct factor. While such a subindex did not improve the performance of the four-component SES index in grades 8 and 12 (when included as an optional fifth component), its inclusion would make a small contribution to variance explained at grade 4, increasing R^2 from 0.21 to 0.23. However, the subindex of technology items is not a conceptually equivalent replacement for the parental education component of SES. Furthermore, the technology subitems are likely to experience changes over the short to medium term given the rapid changes that occur in technology (e.g., see the above analysis of grade 4 smartphone access in 2017 and 2019). Additional investigations and discussion about different options for the grade 4 SES index are needed.

Summary

NAEP is required by law to report on subgroup performance differences, such as those by race/ethnicity, gender, and student family SES background. The current, most prominent proxy variable for SES in NAEP has been National School Lunch Program (NSLP) eligibility, which has been used since 2003. However, concerns have been raised that this indicator may become less valid over time (Cowan et al. 2012), especially after the nationwide implementation of the Community Eligibility Provision (CEP) in the 2014–15 school year under the Healthy, Hunger-Free Kids Act of 2010. Although recent research has pointed to the continued relevance of NSLP status as an SES proxy (Broer, Chen, and Xie 2019; Domina et al. 2018; Xie and Broer 2022a), NSLP eligibility, as a binary variable, has clear limitations both for reporting as well as for being used as an SES control variable in education research.

NCES convened an SES panel (Cowan et al. 2012) that broadly defined SES as one's "access to financial, social, cultural, and human capital resources," while also acknowledging traditional SES definitions from sociology, i.e., the "big three" SES components: parental educational attainment, parental occupational status, and household or family income. Major recommendations from the panel included that NCES develop the "big three" SES indicators or proxies thereof that could form a core SES measure; attempt to develop a composite measure of SES; and explore the inclusion of a school or neighborhood SES component in such an index. To address the recommendations from the SES panel, NCES embarked on a two-pronged strategy.

On the one hand, emphasis was placed on developing, piloting, and validating new SES proxy measures that could complement NSLP eligibility as an income proxy as well as developing proxy measures that could measure the third of the "big three" SES components, occupational prestige. Unfortunately, these new measures did not live up to the expectations that they would make strong contributions to the creation of an SES composite. Without going into detail, in most cases, it was found that students could not respond reliably to these newly developed questions and/or the questionnaire items did not show strong relationships with NAEP performance, nor did they show strong relationships with existing NAEP SES components that do exhibit strong relationships with NAEP performance. While there is still no occupational prestige proxy available, this should not hold up efforts to implement an effective NAEP SES index. The *Contextual Information Framework for the National Assessment of Educational Progress* (NAGB 2013) also stated pragmatically that "although NAEP may never be able to produce a full composite of SES, based on family income, education, and occupation, efforts should be accelerated to develop and use improved measures of socioeconomic status, including an SES index."

On the other hand, NCES encouraged research aimed at exploring the possibility of forming a NAEP SES index with core components long collected by NAEP and validating the utility for such an index. For this purpose, NCES and AIR also developed criteria for evaluating how a possible NAEP SES index performs. These criteria are (1) How well does the NAEP SES index explain NAEP performance? (2) How well does the NAEP SES index account for achievement gaps in NAEP? and (3) Does this index function similarly for major racial/ethnic subgroups?

The NAEP SES index for grades 8 and 12 proposed here is composed of four components that have been collected since 2003: (1) number of books at home; (2) student's eligibility for the National School Lunch Program (NSLP); (3) percentage of students eligible for NSLP at the school the student is attending; and (4) highest level of education of either parent. In grade 4, where no information on parental education is collected from students, the index has only three components. The addition of the school SES component is a response to the NCES SES panel's recommendation to consider adding such a component to an SES composite.

The analyses reported above show that the proposed NAEP SES index performs well against the established criteria: it explains a relatively large amount of performance variation in NAEP (e.g., $R^2 = 0.24$ for grade 8 mathematics), which means that it captures aspects of SES that are relevant to educational outcomes. It also performs much better than NSLP alone in explaining variation in NAEP performance and better than SES indices used in other large-scale surveys administered in the United States (TIMSS, PISA, and HSLS:09). It also accounts for a larger share of the score gap (educational outcome gap) between White students and Black, Hispanic, and Native American/Alaskan Native students than NSLP alone. In the case of the score gap between White and Hispanic students, English learner status and the proposed NAEP SES index almost fully account for the differences in educational outcomes between the two student groups. Moreover, we showed that while the distribution of the NAEP SES index varies strongly by student racial/ethnic subgroup (reflecting longstanding inequities in access to resources that are related to educational outcomes), the proposed NAEP SES index nevertheless performs similarly within each major racial/ethnic subgroup; that is, higher NAEP SES index scores are associated with higher average NAEP scores.

While we believe that the proposed NAEP SES index functions well and presents an opportunity for enhanced NAEP reporting of SES, we understand that there are certain challenges that would need to be addressed for operational implementation of such an index.

One challenge is the robustness of the index, i.e., whether the changes in the underlying components are small enough to enable useful administration-to-administration reporting of score changes by SES levels. We believe that this should not be a problem, but we are also mindful that seen over longer time horizons, the SES construct measured by the SES index can change. For instance, it was likely somewhat different in 2003 compared to 2019 (even though the same four components were available in both years) and might again change between now and 10 years later. To account for changes over longer time periods, we suggest reporting changes in NAEP performance not by SES index points or SES levels based on those index points at a certain point in time, but through a comparison of relative SES levels. For example, one can compare the score differences (or educational outcome gap) between students at the 25th percentile distribution of SES (lower SES) and students at the 75th percentile distribution (higher SES) and track this gap over time, as Bai, Straus, and Broer (2021) have shown. This establishment of relative SES levels also makes comparisons between states with very different SES profiles easier and more pertinent to their specific realities.

Another challenge is how to deal with missing values in NAEP SES index components. We have shown in previous research that imputation of missing SES components appears to be working well (Bai, Straus, and Broer 2021; Broer, Xie, and Bohrnstedt 2017; Xie and Broer 2017), but for operational implementation, additional validation about the best imputation approach to use will still need to be determined. Moreover, decisions will need to be made about whether to include multiple imputed SES values (akin to the multiple plausible values in NAEP) in NAEP data files, which would be a first for a contextual variable.

While we found that the NAEP SES index for grade 4 also works well, the fact that it has one fewer component means that a decision will need to be made to either go ahead with a three-component index, include a fourth component, or not report results by an SES index at grade 4.

While these issues will require further work, discussion, and decisions that will involve trade-offs, we believe that any remaining challenges are solvable and that the NAEP SES index proposed here would be a very useful addition to helping NCES and the NAEP program in its continuous efforts to further contextualize NAEP performance and understand the circumstances associated with varying educational performance.

References

Avvisati, F. (2020). The Measure of Socio-Economic Status in PISA: A Review and Some Suggested Improvements. *Large-scale Assessments in Education*, *8*(1), 1–37.

Bai, Y., Straus, S., and Broer, M. (2021). *U.S. National and State Trends in Educational Inequality Due to Socioeconomic Status: Evidence from the 2003–17 NAEP* (AIR-NAEP Working Paper 2021–01). American Institutes for Research.

Broer, M., Bai, Y., and Fonseca, F. (2019). *Socioeconomic Inequality and Educational Outcomes: Evidence from Twenty Years of TIMSS* (p. 83). Springer Nature.

Broer, M., Chen, J., and Xie, Q. (2019, April 5–9). NSLP Eligibility and Its Relationship with NAEP Performance on the National and State Level. In G. Bohrnstedt (Chair), *Community Eligibility and National School Lunch Program Status. A View from NAEP*. Meeting of the American Educational Research Association, Toronto, Canada.

Broer M., and Xie, Q. (2018). Presentation to the NAEP Design and Analysis Committee on a Potential NAEP SES Index (Internal Advisory Committee Meeting).

Broer, M., Xie, Q., and Bohrnstedt, G. W. (2017, April 27-May 1). Construction and Validation of a New Proxy Socioeconomic Status Measure for NAEP Grade 12. In. J.H. Tang (Chair), *The Development and Applications of Alternative Student Socioeconomic Status Measures*. Symposium conducted at the meeting of the American Educational Research Association, San Antonio, TX.

Burg, S., Stephens, M., Malley, L., and Fonseca, F. (2022, October). *Synthesizing NAEP and International Large-Scale Assessment Score Trends: A Pattern of Diverging Performance*. American Institutes for Research. https://www.air.org/sites/default/files/2022-10/Research-Note-Score-Trends-NAEP-ILSAs-Oct-2022.pdf

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, *84*(3), 517–544.

Choi, S., and Byun, S. Y. (2022). Socioeconomic Inequality in Non-Cognitive Skills Among Young Adolescents in Korea: Gaps, Trends, and Heterogeneity. https://doi.org/10.31235/osf.io/gbm5j

Cowan, C. D., Hauser, R. M., Levin, H. M., Beale Spencer, M., and Chapman, C. (2012). *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*. https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Domina, T., Pharris-Ciurej, N., Penner, A. M., Penner, E. K., Brummet, Q., Porter, S. R., and Sanabria, T. (2018). Is Free and Reduced-Price Lunch a Valid Measure of Educational Disadvantage? *Educational Researcher, 47*(9), 539–555. https://doi.org/10.3102/0013189X18797609

Eriksson, K., Lindvall, J., Helenius, O., and Ryve, A. (2021). Socioeconomic Status as a Multidimensional Predictor of Student Achievement in 77 Societies. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.731634

Geverdt, D. (2018). *Education Demographic and Geographic Estimates Program (EDGE): School Neighborhood Poverty Estimates - Documentation* (NCES 2018-027). U.S. Department of Education. National Center for Education Statistics. https://nces.ed.gov/programs/edge/docs/EDGE_SIDE_PUBSCH_FILEDOC.pdf

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M., and Woessmann, L. (2022). Long-Run Trends in the U.S. SES-Achievement Gap. *Education Finance and Policy, 17*(4), 608–640.

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., and Sterne, J. A. (2014). Joint Modelling Rationale for Chained Equations. *BMC Medical Research Methodology, 14*(1), 28.

Ikoma, S., Bai, Y., and Broer, M. (2017). *Investigating Student vs. Parent-Reported Parental Education Level in the NAEP-HSLS Overlap Sample* (Unpublished research brief prepared for the National Center for Education Statistics). American Institutes for Research.

Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., Rogers, J. E., Kitmitto, S., and Leinwand, S. (2014). *High School Longitudinal Study of 2009 (HSLS:09) Base Year to First Follow-Up Data File Documentation* (NCES 2014-361). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Jewsbury, P. A., Jerry, L., Freund, D. S., and Jia, Y. (2016). *2015 NAEP/ECLS-K Special Study* (Deliverable to the National Center for Education Statistics). Educational Testing Service.

Jewsbury, P. A., Jerry, L., Lu, R., Freund, D. S., Whorton, R. P., Eckerly, C. A., Jia, Y., and Bertling, J. P. (2018). *SES Psychometric Evaluations*. (Deliverable to the National Center for Education Statistics). Educational Testing Service.

Lareau, A. (2011). *Unequal Childhoods: Class, Race, and Family Life*. University of California Press.

Liu, J., Peng, P., and Luo, L. (2020). The Relation Between Family Socioeconomic Status and Academic Achievement in China: A Meta-Analysis. *Educational Psychology Review*, *32*(1), 49–76.

Marks, G. N., and Pokropek, A. (2019). Family Income Effects on Mathematics Achievement: Their Relative Magnitude and Causal Pathways. *Oxford Review of Education*, *45*(6), 769–785.

Martin, M. O., Mullis, I. V. S., Foy, P., and Arora, A. (2012). Creating and Interpreting the TIMSS and PIRLS 2011 Context Questionnaire Scales. In M. O. Martin and I. V. S. Mullis (Eds.), *Methods and Procedures in TIMSS and PIRLS 2011*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Musu-Gillette, L. (2016). Challenges, Changes, and Current Practices for Measuring Student Socioeconomic Status. *National Center for Education Statistics Blog*. https://nces.ed.gov/blogs/nces/post/challenges-changes-and-current-practices-for-measuring-student-socioeconomic-status

National Assessment Governing Board (NAGB). (2013). *Contextual Information Framework for the National Assessment of Educational Progress*. National Assessment Governing Board.

Ogut, B., Bohrnstedt, G., and Broer, M. (2016, April 1–11). Investigating SES Using the NAEP-HSLS Overlap Sample. In J. Kobrin (Chair), *Validating "Noncognitive"/Nontraditional Constructs*. Presented at the 2016 meeting of the National Council on Measurement in Education, Washington, DC.

Organization for Economic Cooperation and Development (OECD). (2019). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. https://doi.org/10.1787/acd78851-en

Putnam, R. (2016). *Our Kids: The American Dream in Crisis*. Simon and Schuster.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. *Whither Opportunity*, *1*(1), 91–116.

Reardon, S. F. (2021). *The Economic Achievement Gap in the US, 1960–2020: Reconciling Recent Empirical Findings* (CEPA Working Paper No. 21-09). Stanford Center for Education Policy Analysis.

Rubin, D. B. (Ed.). (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley. https://doi.org/10.1002/9780470316696

Shipan, R., and Broer, M. (2022, April 21–26). Disadvantage and Devices: Changes in Technology Access by Socioeconomic Status for 2017–2019 NAEP. Presented at the 2022 American Educational Research Association Conference, San Diego, CA. https://doi.org/10.3102/1887935

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453.

Stanley, J. C., and Wang, M. D. (1970). Differential Weighting: A Survey of Methods and Empirical Studies. *Review of Educational Research, 40*(5), 663–705.

Tang, J. H. (2017, April 27–May 1). Examining the Validity and Reliability of Alternative Student-Level Socioeconomic Status Measures. In J. H. Tang (Chair), *The Development and Applications of Alternative Student Socioeconomic Status Measures*. Symposium conducted at the meeting of the American Educational Research Association, San Antonio, TX.

van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3). https://doi.org/10.18637/jss.v045.i03

Xie, Q. (2019, April 5–9). Examining the Reliability of Student Eligibility for NSLP in NAEP with American Community Survey Data. In G. Bohrnstedt (Chair), *Community Eligibility and National School Lunch Program Status. A View from NAEP*. Symposium conducted at the meeting of the American Educational Research Association, Toronto, Canada.

Xie, Q., and Broer, M. (2017, April 27–May 1). Exploring the Construction of a Proxy Socioeconomic Status Index for NAEP Grade 4 Assessments. In. J. H. Tang (Chair), *The Development and Applications of Alternative Student Socioeconomic Status Measures*. Symposium conducted at the annual meeting of the American Educational Research Association, San Antonio, TX.

Xie, Q., and Broer, M. (2022a). *An Analysis of the Validity of Student Eligibility Status for NSLP as a Proxy Variable for Socio-economic Status in Reporting NAEP Results After the Implementation of Community Eligibility Provision* (Unpublished research brief prepared for the National Center for Education Statistics). American Institutes for Research.

Xie, Q., and Broer, M. (2022b, April 21–26). School Poverty – Comparing a Measure Based on Census Data with School NSLP Percentage. In B. J. Park (Chair). Symposium paper accepted to the meeting of the 2022 American Educational Research Association, San Diego, CA.

Yavuz, S., Bai, Y., Zheng, X., and Broer, M. (2022). *Multiple Imputation Approach to Missingness at the School Level: A Comparison Study* (Unpublished research brief prepared for the National Center for Education Statistics). American Institutes for Research.

Yin, L., and Fishbein, B. (2020). Creating and Interpreting the TIMSS 2019 Context Questionnaire Scales. In M. O. Martin, M. von Davier, and I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 16.1–16.331). Retrieved from https://timssandpirls.bc.edu/timss2019/methods/chapter-16.html

Appendix A. Description of the Analytic Sample

For each NAEP assessment cycle, a sample of students in designated grades within both public and private schools throughout the United States (and sometimes specified territories and possessions) is selected for assessment. For our analysis, we performed the following procedure to obtain the analytic sample for each grade and subject. Here we use the NAEP 2019 grade 8 mathematics assessment data as an example for illustration:

- Restrict the analysis to the national reporting sample (RPTSAMP = 1, *N* = 147,430).[16]
- Restrict the analysis to the 50 states and the District of Columbia and exclude students from U.S. territories (e.g., Virgin Islands, Guam), Bureau of Indian Education (BIE) funded schools, and schools oversea (DoDEA/DoDDS schools) (*N* = 144,860).
- Restrict the analysis to the national public school sample (*N* = 142,200).
- Restrict the analysis to the states that have valid questionnaire information for SES components. More specifically, six states that do not have information on parental education or number of books at home were excluded from the analysis. These states are Alaska, Colorado, Montana, New Hampshire, South Dakota, and Utah (*N* = 127,920).
- Restrict the analysis to the cases with no missing values on any of the SES components (*N* = 108,930).

[16] In accordance with Institute of Education Sciences (IES) statistical standards, all unweighted sample size numbers in this report are rounded to the nearest 10.

Appendix B. The Handling of Missing Data in School NSLP

In this study, we used data from the Common Core of Data (CCD) for public schools in the 2018–19 school year to mitigate the severe missingness of school NSLP in NAEP data. To be specific, we calculated the percentage of students eligible for NSLP at each school by utilizing the total number of students enrolled in a school and the number of students eligible for free or reduced-price lunch in that school. Here are the steps taken to create a new variable with lower percentage of missingness on school NSLP using CCD and NAEP data.

First, we limited our analyses to all public schools in the 50 states and the District of Columbia in both the CCD and NAEP. We merged the CCD data with the NAEP data using the unique school ID assigned by NCES. Out of the 6,560 schools in NAEP, 6,430 schools found a match in the CCD data (Table B.1).

Table B.1. Number of schools and number of students matched between 2019 grade 8 NAEP mathematics and the 2018–19 CCD

Match result	Data source	Number of schools	Number of students
Not matched	NAEP	140	2,890
	CCD	93,180	N/A
Matched	NAEP and CCD	6,430	139,320
Total	NAEP	6,560	142,200

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment; and 2018–19 Common Core of Data.

Table B.2 and Table B.3 provide the distributions of school NSLP generated from the CCD data and from the NAEP data, respectively. Given the much smaller missing rate, we treated the CCD distribution as the base for missing data manipulation.

Table B.2. Frequency and percentage distribution of school NSLP component values in the 2018–19 CCD

School NSLP component value	Frequency	Percent	Cumulative percent
0 (76–100% eligible)	31,710	22.3	22.3
1 (51–75% eligible)	32,070	22.6	44.9
2 (26–50% eligible)	39,880	28.1	72.9
3 (0–25% eligible)	22,050	15.5	88.4
Missing	16,500	11.6	100.0
Total	142,200	100.0	100.0

NOTE: The school National School Lunch Program (NSLP) component values were coded as follows: schools with 76–100% of students eligible for NSLP were coded as 0; schools with 51–75% eligible were coded as 1; schools with 26–50% eligible were coded as 2; and schools with 0–25% eligible were coded as 3. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2018–19 Common Core of Data.

Table B.3. Frequency and percentage distribution of school NSLP component values in 2019 grade 8 NAEP mathematics

School NSLP component value	Frequency	Percent	Cumulative percent
0 (76–100% eligible)	8,540	6.0	6.0
1 (51–75% eligible)	19,570	13.8	19.8
2 (26–50% eligible)	34,680	24.4	44.2
3 (0–25% eligible)	22,560	15.9	60.0
Missing	56,850	40.0	100.0
Total	142,200	100.0	100.0

NOTE: The school National School Lunch Program (NSLP) component values in this table are based on the reported school-level NSLP variable in NAEP (C051651) and were coded as follows: schools with 76–100% of students eligible for NSLP were coded as 0; schools with 51–75% eligible were coded as 1; schools with 26–50% eligible were coded as 2; and schools with 0–25% eligible were coded as 3. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Next, we used the non-missing values from the NAEP data to replace the missing values in the CCD (Table B.4).

Table B.4. Frequency and percentage distribution of school NSLP component values based on the 2018–19 CCD and 2019 grade 8 NAEP mathematics

School NSLP component value	Frequency	Percent	Cumulative percent
0 (76–100% eligible)	32,340	22.7	22.7
1 (51–75% eligible)	33,270	23.4	46.1
2 (26–50% eligible)	42,380	29.8	75.9
3 (0–25% eligible)	24,820	17.5	93.4
Missing	9,400	6.6	100.0
Total	142,200	100.0	100.0

NOTE: The school National School Lunch Program (NSLP) component values were coded as follows: schools with 76–100% of students eligible for NSLP were coded as 0; schools with 51–75% eligible were coded as 1; schools with 26–50% eligible were coded as 2; and schools with 0–25% eligible were coded as 3. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment; and 2018–19 Common Core of Data.

Then, for each school, we calculated the percentage of NSLP-eligible students among all sampled students in that school and used the corresponding score from 0 to 3 (see Table 1) for that percentage to replace the remaining missing values (Table B.5).

Table B.5. Frequency and percentage distribution of school NSLP component values based on the 2018–19 CCD and 2019 grade 8 NAEP mathematics

School NSLP component value	Frequency	Percent	Cumulative percent
0 (76–100% eligible)	36,480	25.7	25.7
1 (51–75% eligible)	34,860	24.5	50.2
2 (26–50% eligible)	44,830	31.5	81.7
3 (0–25% eligible)	25,880	18.2	99.9
Missing	160	0.1	100.0
Total	142,200	100.0	100.0

NOTE: The school National School Lunch Program (NSLP) component values were coded as follows: schools with 76–100% of students eligible for NSLP were coded as 0; schools with 51–75% eligible were coded as 1; schools with 26–50% eligible were coded as 2; and schools with 0–25% eligible were coded as 3. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment; and 2018–19 Common Core of Data.

Appendix C. Correlation Matrix of SES Index Components and NAEP Performance

Appendix C contains two tables, for grade 8 and grade 4 NAEP mathematics, which show the correlations among mathematics achievement and the NAEP SES index components: number of books at home, student's NSLP eligibility, percentage of NSLP-eligible students at the student's school, and highest level of education of either parent (in grade 8).

Table C.1. Matrix of correlations for grade 8 NAEP mathematics achievement, number of books at home, student NSLP eligibility, highest level of parental education, and school NSLP: 2019

Variables	Mathematics achievement	Number of books at home	Student NSLP eligibility	Percentage of NSLP-eligible students at school	Highest level of parental education
Mathematics achievement	1.00	0.38	0.38	0.37	0.29
Number of books at home	0.38	1.00	0.32	0.31	0.33
Student NSLP eligibility	0.38	0.32	1.00	0.51	0.38
Percentage of NSLP-eligible students at school	0.34	0.31	0.51	1.00	0.33
Highest level of parental education	0.29	0.33	0.38	0.33	1.00

NOTE: The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." The school NSLP variable is derived from CCD and NAEP data; for more detail, see Appendix B.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Table C.2. Matrix of correlations for grade 4 NAEP mathematics achievement, number of books at home, student NSLP eligibility, and school NSLP: 2019

Variables	Mathematics achievement	Number of books at home	Student NSLP eligibility	Percentage of NSLP-eligible students at school
Mathematics achievement	1.00	0.29	0.38	0.37
Number of books at home	0.29	1.00	0.29	0.29
Student NSLP eligibility	0.38	0.29	1.00	0.54
Percentage of NSLP-eligible students at school	0.37	0.29	0.54	1.00

NOTE: Parents' highest level of education is not collected in grade 4. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." The school NSLP variable is derived from CCD and NAEP data; for more detail, see Appendix B.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4 Mathematics Assessment.

Appendix D. Relationship of NSLP Eligibility and the SES Index with NAEP Performance

Table D.1 presents the results of binary regressions that were conducted to examine how well the NAEP SES index explains students' performance compared to NSLP eligibility alone, for grade 4, grade 8, and grade 12 NAEP mathematics in 2019.

Table D.1. Results of binary regression of grade 4, grade 8, and grade 12 NAEP mathematics performance on National School Lunch Program (NSLP) eligibility and the NAEP SES index: 2019

	Grade 4 Mathematics		Grade 8 Mathematics		Grade 12 Mathematics	
NAEP 2019	NSLP eligibility	SES index	NSLP eligibility	SES index	NSLP eligibility	SES index
Coefficient	24.0***	5.2***	29.8***	5.8***	23.9***	5.0***
	(0.34)	(0.06)	(0.45)	(0.06)	(0.67)	(0.09)
Constant	229.5***	217.2***	268.1***	243.5***	137.2***	115.4***
	(0.21)	(0.32)	(0.32)	(0.50)	(0.45)	(0.69)
R^2	0.14	0.21	0.14	0.24	0.11	0.22

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
NOTE: NSLP = National School Lunch Program. The NAEP mathematics scale ranges from 0 to 500 at grades 4 and 8 and from 0 to 300 at grade 12. The information about student NSLP eligibility is based on available school records. If school records were not available, the student was classified as "Information not available." Standard errors are shown in parentheses. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment.

Appendix E. The Achievement Gap Between White Students and Other Racial Groups

Appendix E contains the results of analytic models examining achievement gaps between different racial/ethnic groups for grade 8 and grade 4 NAEP mathematics, before and after controlling for LEP and either NSLP or the NAEP SES index. Model 1, or the null model, examined the achievement gaps between racial/ethnic groups without controlling for any variables. Model 2 added LEP as a control variable. Model 3 added students' eligibility for NSLP as an additional control variable, while Model 4 added students' scores on the NAEP SES index instead of NSLP.

Table E.1. Achievement gaps between White students and other racial/ethnic groups for grade 8 NAEP mathematics, by model: 2019

	Model 1			Model 2 (Model 1 + LEP)			Model 3 (Model 2 + NSLP)			Model 4 (Model 2 + SES)		
Grade 8 Mathematics	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Reference group: White												
Black	-32.2	***	(0.60)	-32.0	***	(0.60)	-23.1	***	(0.54)	-17.2	***	(0.49)
Hispanic	-22.4	***	(0.59)	-15.9	***	(0.60)	-7.9	***	(0.64)	0.1		(0.60)
Asian/Pacific Islander	20.3	***	(1.32)	23.2	***	(1.22)	23.9	***	(1.12)	23.2	***	(1.04)
American Indian/ Alaska Native	-26.3	***	(1.79)	-24.7	***	(1.87)	-16.6	***	(1.78)	-10.9	***	(1.63)
Two or more races	-6.1	***	(0.98)	-6.1	***	(0.98)	-3.1	***	(0.90)	-2.0	**	(0.90)
LEP				-37.3	***	(1.04)	-32.8	***	(1.07)	-27.6	***	(1.07)
Not eligible for NSLP							21.0	***	(0.43)			
SES index (12 points)										4.8	***	(0.06)
Constant	293.3	***	(0.32)	293.6	***	(0.32)	278.8	***	(0.37)	253.5	***	(0.55)
N	108,930			108,930			108,930			108,930		
R^2	0.14			0.18			0.24			0.30		

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
NOTE: LEP = limited English proficiency. NSLP = National School Lunch Program. The NAEP mathematics scale ranges from 0 to 500 at grade 8. The SES index ranges from 0 to 12 points at grade 8. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Table E.2. Achievement gaps between White students and other racial/ethnic groups for grade 4 NAEP mathematics, by model: 2019

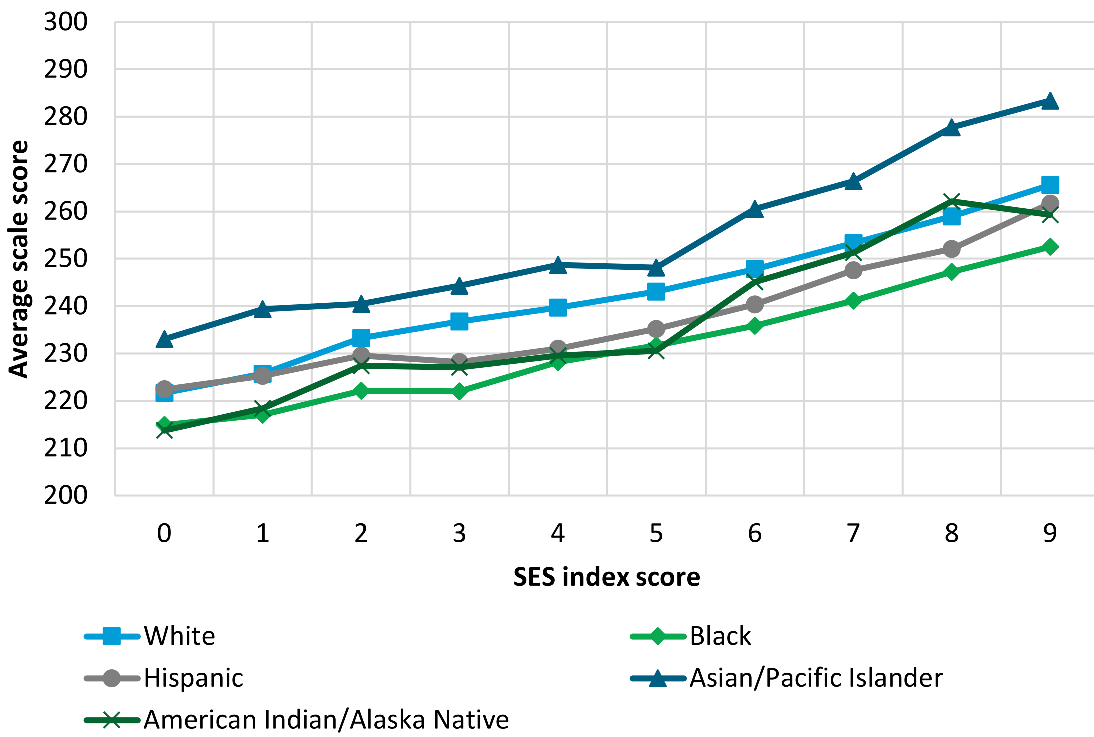
	Model 1			Model 2 (Model 1 + LEP)			Model 3 (Model 2 + NSLP)			Model 4 (Model 2 + SES)		
Grade 4 Mathematics	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Reference group: White												
Black	-24.5	***	(0.45)	-24.2	***	(0.46)	-17.0	***	(0.44)	-12.3	***	(0.46)
Hispanic	-17.6	***	(0.43)	-10.2	***	(0.43)	-4.3	***	(0.44)	-0.8		(0.45)
Asian/Pacific Islander	13.0	***	(1.07)	17.2	***	(1.04)	16.4	***	(0.89)	16.5	***	(0.81)
American Indian/ Alaska Native	-18.1	***	(1.33)	-16.8	***	(1.31)	-10.7	***	(1.21)	-7.2	***	(1.17)
Two or more races	-5.3	***	(0.66)	-5.3	***	(0.66)	-2.5	***	(0.61)	-1.6	**	(0.62)
LEP				-21.2	***	(0.64)	-18.2	***	(0.63)	-16.1	***	(0.63)
Not eligible for NSLP							17.1	***	(0.29)			
SES index (9 points)										4.0	***	(0.06)
Constant	248.8	***	(0.26)	249.1	***	(0.26)	238.1	***	(0.30)	225.6	***	(0.40)
N	125,900			125,880			125,880			125,880		
R^2	0.12			0.16			0.22			0.25		

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
NOTE: LEP = limited English proficiency. NSLP = National School Lunch Program. The NAEP mathematics scale ranges from 0 to 500 at grade 4. Parents' highest level of education is not collected in grade 4; therefore, the grade 4 SES index has only three components and ranges from 0 to 9. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. The information about student eligibility for the National School Lunch Program (NSLP) is based on available school records. If school records were not available, the student was classified as "Information not available." Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4 Mathematics Assessment.

Appendix F. Average Achievement by SES Index Score and Race/Ethnicity

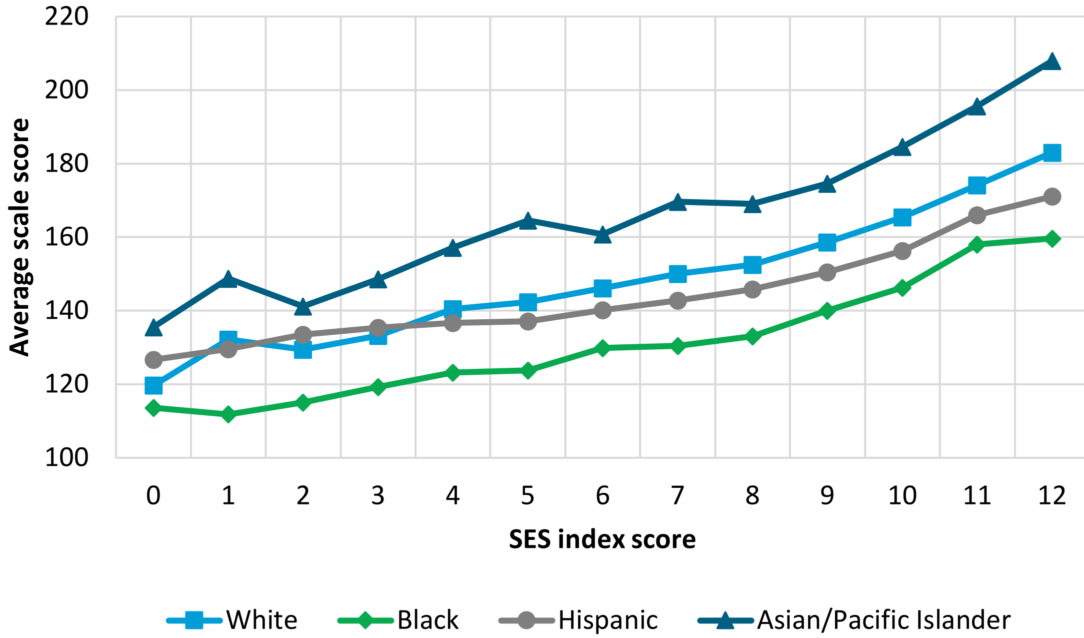
Figure F.1 and Figure F.2 present the average achievement score by each SES index score for each racial group in grade 4 and grade 12 NAEP mathematics, respectively, in 2019. The box plot in figure F.3 presents a more detailed view of the relationship between average scale scores and SES index scores in grade 8 NAEP mathematics in 2019.

Figure F.1. Average scale score for grade 4 NAEP mathematics, by SES index score and race/ethnicity: 2019



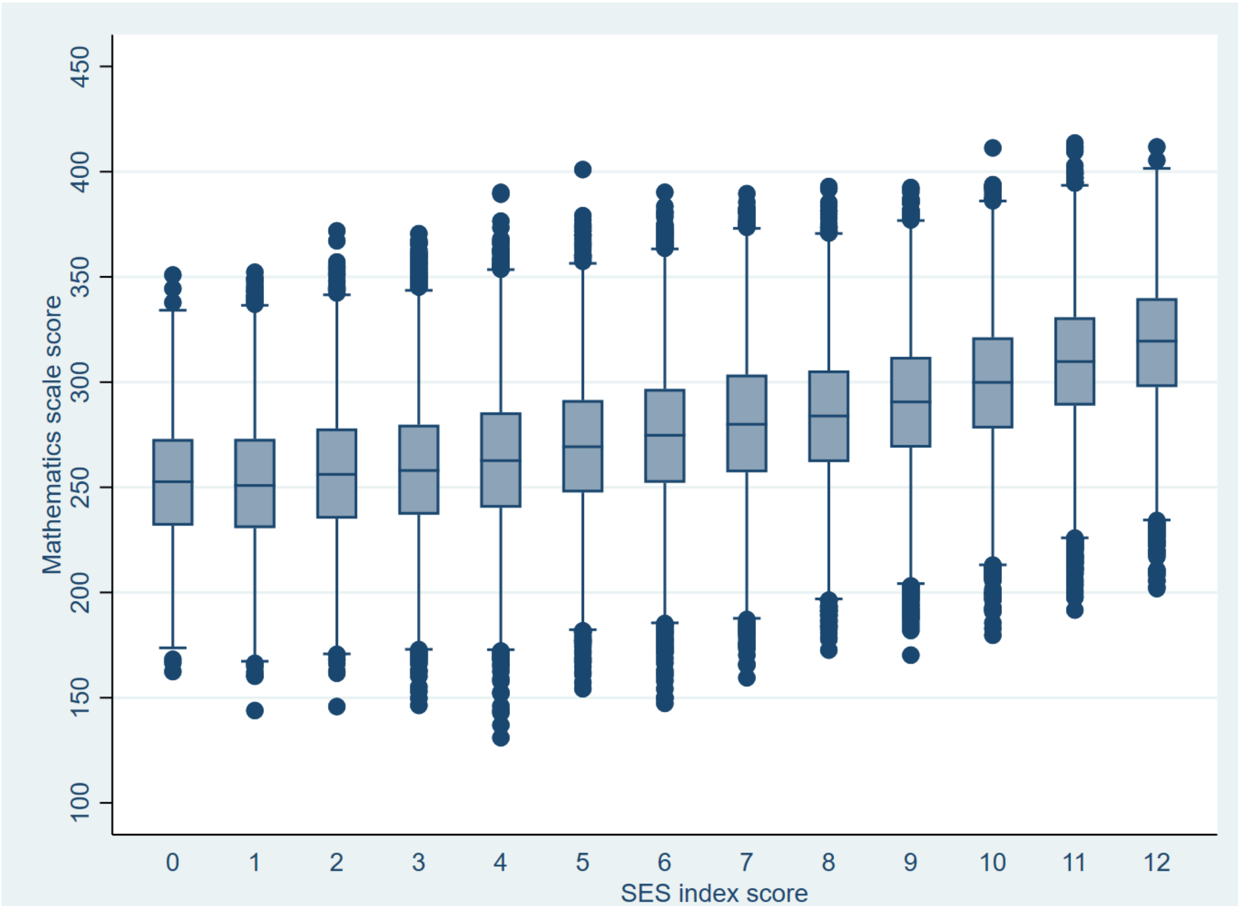
NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 4. The SES index ranges from 0 to 9 points at grade 4. Parents' highest level of education is not collected in grade 4; therefore, the grade 4 SES index has only three components. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. Students of Two or More Races assessed in grade 4 NAEP mathematics are not included in this figure due to small sample sizes. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4 Mathematics Assessment.

Figure F.2. Average scale score for grade 12 NAEP mathematics, by SES index score and race/ethnicity: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 300 at grade 12. The SES index ranges from 0 to 12 points at grade 12. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin. American Indian/Alaska Native students and students of Two or More Races assessed in grade 12 NAEP mathematics are not included in this figure due to small sample sizes. Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 12 Mathematics Assessment.

Figure F.3. Box plot of student scale score for grade 8 NAEP mathematics, by SES index score: 2019



NOTE: The NAEP mathematics scale ranges from 0 to 500 at grade 8. The NAEP mathematics scale score here is the mean of 20 plausible values. The SES index ranges from 0 to 12 points at grade 8.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 8 Mathematics Assessment.

Appendix G. SES-related Indices in Other Major Student Assessments Administered in the United States: TIMSS, PISA, and HSLS:09

Trends in International Mathematics and Science Study (TIMSS)

The current home educational resources (HER) scale used in TIMSS 2019 (Yin and Fishbein 2020) was created based on students' reports regarding the availability of the three resources listed below:

- *Number of books in the home:* This information comes from the student questionnaire asking how many books students have at home. There are five categories: (1) 0 to 10 books; (2) 11 to 25 books; (3) 26 to 100 books; (4) 101 to 200 books; and (5) more than 200 books.
- *Number of home study supports:* This information is derived from questions asking students whether they have the following study supports at home: (1) None; (2) Internet connection or own room; and (3) Both internet connection and own room.
- *Highest level of education of either parent:* This information is derived from both the father's and mother's highest educational levels. The categories of the source variables were grouped into five levels: (1) Finished some primary or lower secondary or did not go to school; (2) Finished lower secondary; (3) Finished upper secondary; (4) Finished postsecondary education; and (5) Finished university or higher.

Though initially a categorically defined measure based on the levels of its components (low, middle, high), the HER index has been constructed since 2011 using IRT scaling methodology (Martin et al. 2012).

Program for International Student Assessment (PISA)

The economic, social, and cultural status (ESCS) index in PISA 2018 (OECD 2019) is a composite score derived from the following three indicators, which are based on student reports:

- *Parents' highest level of education:* This indicator represents the highest ISCED classification of the levels of schooling attained by the student's parents, based on student reports and mapped to (approximate) years of schooling for each participating country.
- *Parents' highest occupational status:* This indicator represents the highest occupation status of the student's parents. Occupation variables reported by students are quantified based on the International Standard Classification of Occupations (ISCO-08).
- *Index of home possessions:* This index represents the availability of various household items at home, based on 25 items asked of students in the background questionnaire.

These items include home educational resources (e.g., a desk to study at), items related to "classical" culture (e.g., books of poetry), and other indicators of family wealth (e.g., a guest room), in addition to the number of books in the home.

In PISA 2018, ESCS is computed by attributing equal weight to the three components, which were standardized across all participating countries and economies, with each country/economy contributing equally (OECD 2019). In previous cycles, ESCS was derived from a principal component analysis of standardized variables, taking the factor scores for the first principal component as measures of the ESCS index. The final ESCS variable was transformed to have an OECD mean of 0 and a standard deviation of 1 across equally weighted OECD countries. For students with one missing component, values were imputed with predicted values plus a random component based on a regression on the other two variables. ESCS was not computed for students with more than one missing component.

High School Longitudinal Study of 2009 (HSLS:09)

The SES index in HSLS:09 is constructed as a function of the following five variables obtained through the parent/guardian questionnaire:

- *Highest education among the two parents/guardians* in a two-parent family home, or the education of the sole parent/guardian in a single-parent home,
- *Education level of the second parent/guardian* in a two-parent family home,
- *Highest occupational prestige score among the two parents/guardians* in a two-parent/guardian family, or the prestige score of the sole parent/guardian in a single-parent home,
- *Occupational prestige score of the second parent/guardian* in a two-parent family home, and
- *Family income*.

Each component was standardized and combined into an SES composite score. For the full methodology, refer to Ingels et al. (2014).

Table G.1. SES indices and components in TIMSS, PISA, and HSLS:09

Assessment	Index	Components
TIMSS	Home educational resources (HER)	Highest parental education level
		Number of books in the home
		Number of home study supports
PISA	Economic, social, and cultural status (ESCS)	Highest parental education level
		Highest parental occupation status
		Index of home possessions (educational resources, cultural items, indicators of family wealth, and books in the home)
HSLS:09	SES index	Highest parental education level
		Education level of second parent/guardian
		Highest parental occupational prestige score
		Occupational prestige score of second parent/guardian
		Family income

SOURCE: Ingels et al. 2014; OECD 2019; Yin and Fishbein 2020.

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit AIR.ORG.

American Institutes for Research®
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
202.403.5000 | **AIR.ORG**