RESEARCH   **Open Access**

The relationship between mathematics self-efficacy and mathematics achievement: multilevel analysis with NAEP 2019

Yao Yang[1]* , Yukiko Maeda[2] and Marcia Gentry[2]

*Correspondence:
yao.yang@yale.edu

[1] Poorvu Center for Teaching and Learning, Yale University, 301 York St, New Haven, CT 06511, USA
[2] Department of Educational Studies, Purdue University, 100 N University St, West Lafayette, IN 47907, USA

Abstract

Background: This empirical study aims to investigate the association between mathematics self-efficacy and mathematics achievement gaps among students in Grades 4, 8, and 12, utilizing data from the 2019 National Assessment of Educational Progress (NAEP). The study also considers student-level (e.g., mathematics self-efficacy, gender, race/ethnicity) and school-level (e.g., school location, proportion of underrepresented students) demographics to provide a comprehensive analysis of the factors contributing to mathematics achievement gaps.

Methods: A two-level cross-sectional multilevel modeling approach was employed to analyze the variance in mathematics achievement, partitioning it into within- and between-school components. This approach allowed for an examination of association between mathematics self-efficacy and achievement gaps while considering various student and school characteristics. The rationale behind this methodology lies in its ability to capture the hierarchical nature of educational data and provide a nuanced understanding of the factors associated with mathematics achievement.

Results: The analysis of the NAEP data revealed substantial variability in mathematics achievement across schools in the United States at all grade levels. Furthermore, mathematics self-efficacy emerged as a robust predictor of students' mathematics achievement, exhibiting significant effect sizes for Grades 4, 8, and 12. Remarkably, when students' mathematics self-efficacy was held constant, the mathematics achievement gaps among different student subgroups by gender, race/ethnicity, ELL, IEP, NSLP status narrowed, highlighting the importance of self-efficacy in addressing these disparities. The study also identified the presence of significant school contextual effects, further emphasizing the role of the educational environment in shaping mathematics achievement.

Conclusions: This study underscores the critical role of mathematics self-efficacy in influencing mathematics achievement gaps among students. By acknowledging the association between self-efficacy beliefs and mathematics achievement, policymakers and educators can develop targeted interventions to enhance students' confidence and motivation in mathematics, ultimately promoting equitable educational outcomes. The findings also emphasize the significance of school-level factors, calling for comprehensive approaches that consider both individual and contextual factors in narrowing achievement gaps. The implications of adopting a self-efficacy

perspective to address mathematics achievement gaps extend to educational policy, curriculum development, and instructional practices, ultimately fostering more inclusive and effective mathematics education for all students.

Keywords: NAEP, Mathematics self-efficacy, Mathematics achievement, Achievement gap

Introduction

The disparities in mathematics achievement by student subgroups have been a critical issue in addressing the inequities in mathematics education (Hanushek et al., 2019). In previous achievement gap studies, attention has been given to highlighting persistent achievement gaps by accounting for students' sociodemographic characteristics and school contextual differences (e.g., Lee & Reeves, 2012; Lubienski & Lubienski, 2006; Young et al., 2017). However, caution must be stressed in overemphasizing the associations between students' demographic characteristics and mathematics achievement, because it may reinforce the deficit views toward the underrepresented students when years of efforts to enhance the achievement of underrepresented students do not yield desirable educational outcomes (Garcia & Guerra, 2004). Also, over-emphasis on the effects of demographic markers may take our attention away from possible omitted variables that are associated with achievement, potentially obscuring the broader landscape of influences, including psychological aspects (Wilms et al., 2021). Therefore, we propose that studies exploring achievement gaps should be reframed from the anti-deficit and asset-oriented approach (Harper, 2010) to identify how some malleable factors, such as self-efficacy, are associated with the observed mathematics achievement gaps.

Mathematics self-efficacy is closely associated with students' mathematics achievement. Self-efficacy refers to an individual's beliefs in their capacities to execute courses of action to accomplish and succeed in given tasks (Bandura, 1977). According to social cognitive theory (Bandura, 1986), students' self-efficacy is associated with individual factors, such as goal setting (Schunk & DiBenedetto, 2021), as well as by environmental factors, such as schools where students belong (Bandura, 2001). Studies (e.g., Sakellariou, 2022) have shown a consistent positive correlation between self-efficacy in mathematics and students' mathematics achievement.

Nevertheless, at least three gaps in the current studies on the role of mathematics self-efficacy in enhancing mathematics achievement remain. First, methodologically, previous studies (e.g., Soland & Sandilos, 2021) tend to overlook the potential associations between school environments and mathematics self-efficacy or mathematics achievement; as a result, the broader interpretative value of these findings may be constrained. Second, the relationship between mathematics self-efficacy and achievement has been shown to co-vary when considering some affective variables (e.g., mathematics anxiety, Kalaycıoğlu, 2015; interest, Zhang & Wang, 2020). These variables were often reported to be moderately to highly correlated with self-efficacy (e.g., Hiller et al., 2022). This may result in underappreciation of the extent to which mathematics self-efficacy and other affective variables are associated with mathematics achievement. Finally, while the association between mathematics self-efficacy and achievement may differ by students' developmental stages (Shell et al., 1995), such potential differences by age or grade levels are rarely discussed. Thus, we designed this study to address these gaps in the literature

by providing a better understanding of the association between mathematics self-efficacy and achievement gaps across grades and contexts. More specifically, we used a multilevel modeling (MLM) framework to nest students into schools to explore the association of mathematics self-efficacy with achievement, considering a series of student- and school-level demographics with data from the National Assessment of Educational Progress (NAEP) in Grades 4, 8, and 12.

With the findings from this study about academic self-efficacy, we aim to shift the attention to psychological variables that are malleable and may correlate with narrower achievement gaps among students from different demographic subgroups. Therefore, in this study, we sought to (1) estimate the degree to which mathematics self-efficacy is associated with student mathematics achievement; (2) examine changes in mathematics achievement gaps among student demographic subgroups when accounting for variations in mathematics self-efficacy; and (3) explore if average mathematics self-efficacy at each school is related to their overall school mathematics achievement.

The association between mathematics self-efficacy and mathematics achievement

Research tends to report a positive association between self-efficacy and mathematics achievement. For example, Kalaycıoğlu, (2015) explored the association of mathematics self-efficacy, socioeconomic status (SES), and mathematics anxiety with secondary school students' mathematics achievement in PISA 2012 and found mathematics self-efficacy to be a strong correlational predictor across all countries with medium effect sizes. Similarly, in a study using PISA 2012 in Greece, Pitsia et al., (2017) found that mathematics self-efficacy was a significant correlational factor in predicting middle-school students' mathematics achievement after controlling gender and school mean SES, among a group of non-cognitive factors (e.g., mathematics self-concept, attitudes toward school, intrinsic motivation, and instrumental motivation). Moreover, in a meta-analysis on psychological correlates of academic achievement, Richardson et al., (2012) reported a medium-sized positive correlation between academic self-efficacy and academic achievement. They also found self-efficacy had the strongest correlations with academic achievement among compared to approximately fifty other achievement-related cognitive and non-cognitive variables. Mathematics self-efficacy was associated with a considerable proportion of the variance in mathematics achievement alone, accounting for about 54% in Cheema and Kitsantas, (2014) and 40% in Kitsantas (2011) within U.S. PISA data.

The association of mathematics self-efficacy with mathematics achievement across student subgroups

The literature generally highlights the positive correlation between self-efficacy and academic performance, noting variations in the association among student subgroups. For example, gender has been linked to differences in students' mathematics self-efficacy, particularly during their high school years. Huang's meta-analysis (2013) of 187 studies from elementary to high school ages indicated no statistically significant gender differences in mathematics self-efficacy existed among students either at elementary or middle school, but found such difference in high school with males reporting higher

self-efficacy in mathematics. Conversely, Schwery, (2015) compared fifth- to eighth-grade students and found no significant gender gap in mathematics self-efficacy levels or their mathematics achievement, nor in the strength of the association between mathematics self-efficacy and mathematics achievement. These findings may suggest that the often-reported gender gap in mathematics achievement in high school or beyond may be associated, to some extent, with the reported difference in mathematics self-efficacy between genders (e.g., Cheema & Galluzzo, 2013).

Likewise, a student's status as an English language learner (ELL) has been associated with variations in mathematics self-efficacy. Sandilos et al., (2020) investigated elementary ELLs and reported lower levels of self-efficacy and achievement in mathematics, even when student demographics and working memory were accounted for. They also found no significant association between limited English proficiency and either mathematics self-efficacy or mathematics achievement. In contrast, Soland and Sandilos, (2021) found an association between ELL status and the growth trajectory of mathematics achievement through self-efficacy in middle school students. Variations in findings between these studies may be partially linked to their differing sample compositions in terms of race/ethnicity. Sandilos et al., (2020) focused on a sample predominantly composed of White students (43%) and a small proportion of ELLs (16%), whereas Soland & Sadilos, (2021) examined a longitudinal cohort from a large urban middle school district with a significant representation of Hispanic (70%) and ELL students (30%). Therefore, the association involving ELL status reported in these studies might also relate to other socio-demographic factors within the samples.

Students with learning disabilities are often observed to have lower self-efficacy in mathematics. Larsen and Jang, (2022) noted an association between students with IEP in Grade 6 and lower self-efficacy compared to their non-IEP peers. Additionally, they found that when students with IEPs were placed in classes with inquiry-based teaching, there was an association with improved mathematics achievement. Similarly, Jungert & Anderson, (2013) reported that the lower mathematics self-efficacy among fifth graders with learning disabilities correlated with a history of challenges in mathematics performance and prolonged experience of difficulty with mathematics learning. Their research suggests that instructional practices aimed at enhancing self-efficacy could be associated with supporting achievement among students with special needs.

The literature consistently reports an association between students' SES and their self-efficacy in mathematics, which in turn is correlated with their mathematics achievement. Wiederkehr et al., (2015) found a correlation between lower SES and decreased levels of mathematics self-efficacy among elementary and secondary students, with this group also tending to have lower mathematics achievement compared to their higher SES peers. These associations were statistically significant with moderate to strong effect sizes. Similarly, McConney and Perry, (2010) analyzed PISA 2003 data and found that students with a higher level of self-efficacy were associated with better mathematics performance when controlling SES status. Yet, the used measures of students' SES in these studies are often not consistent, such as using parents' occupations (e.g., Wiederkehr et al., 2015), or constructed SES index with few variables (e.g., McConney & Perry, 2010). Such variability in measuring SES may relate to the observed strength of its association with self-efficacy or achievement.

Finally, by using a large-scale dataset, Kotok, (2017) found that high-achieving African Americans self-reported the highest degree of self-efficacy whereas high-achieving Asian students had the lowest. After controlling for some student- and school-level variables (e.g., family background, peer engagement, math efficacy, mathematics identity, school SES, school types, and school locations), inversely, high-achieving African American students with higher mathematics self-efficacy were associated with lower mathematics achievement compared to their White and Hispanic counterparts. Although Kotok, (2017) only use the sub-sample of high-achieving students in this study, the result partly supports the results by Cheema and Kitsantas, (2014) that reported a small, but significant negative relationship between mathematics self-efficacy and mathematics achievement among Black students in PISA 2003 U.S. sample. Nevertheless, with NSLS: 2009, Andersen and Ward (2014) did not detect any statistically significant differences in the association between mathematics self-efficacy and achievement among Black, Hispanic, and White high-achieving students.

School contexts and their relationship with mathematics achievement

Students develop their academic knowledge through the dynamic interplays between students' self-efficacy and social-ecological factors (Bandura, 2001), and thus, considering school contexts is crucial for a more comprehensive understanding of the association between self-efficacy and mathematics achievement. Nevertheless, findings regarding the significance of school location as related to mathematics achievement vary. Kotok, (2017) did not find a significant association between school location (city, town, rural) and achievement disparities. In contrast, Wu, (2015), using NAEP data from 2000 to 2011 for Native American students at advanced academic levels, reported that regional school location was related to achievement in Grades 4 and 8. With PISA 2000 U.S. data, Williams, (2005) found a marginal rural-town achievement gap in mathematics but a substantial urban-town achievement gap in favor of town schools. Webster and Fisher, (2010) analyzed the TIMSS Australian sample and found that rural school students were associated with lower mathematics scores compared to urban school students, even when considering differences in socioeconomic status.

While Webster and Fisher, (2010) found a higher availability of mathematics resources in rural schools compared to urban schools, the presence of mathematics resources did not show a substantial association with mathematics achievement. It leaves us to wonder if there are other potential variables might play a role in the relationship between school location and school-wide achievement. For example, school SES, reflecting the average of all students' SES in a school (as used by McConney & Perry, 2010), is reported as a significant factor related to a school's overall achievement (*e.g.,* Kotok, 2017; Pitsia et al., 2017), but then the relationships between school SES and mathematics achievement vary by school location as well. For example, Williams, (2005) found that school SES was particularly influential in urban schools, more so than in rural ones, in relation to mathematics achievement. In town schools, a significant correlation was observed between school SES and student mathematics achievement. When school SES was accounted for, the gap in mathematics achievement between rural and town schools diminished, yet the difference between urban and town schools persisted, with town schools having the advantage. McConney and

Perry, (2010) implied that regardless of their mathematics self-efficacy levels, students being part of a high SES school group were associated with higher mathematics achievement compared to those in lower SES schools after controlling students' individual SES status. The strongest association between school SES and mathematics achievement was seen in students with high self-efficacy from less affluent families. Therefore, more thorough investigations of the school's environmental influences are warranted for a better understanding of student mathematics achievement.

Related to this point, to date, there has been little discussion on how access to gifted education programs at schools is associated with students' mathematics achievement. Therefore, in this study, we included the variable that indicates access to the gifted and talented programs at school as an additional school contextual variable. We view the percentage of gifted program participation to represent the rate of access to gifted education by following a talent development paradigm that focuses on increasing students' access to high-quality gifted services to develop their talents in certain areas (Gentry et al., 2021). It is commonly accepted to view that schools with a higher percentage of students attending gifted programs tend to have higher average mathematics achievement because only already high-achieving students are identified to attend gifted programs. Such a view followed the traditional gifted child paradigm that assumes gifted children are born with high potentials (e.g., IQ, Terman, 1925) and the goal of gifted education is to make the fullest use of these potentials (Dai & Chen, 2013). Researchers (e.g., Gentry et al., 2021) have criticized this paradigm for substantially excluding a broad range of students with gifts and talents in domain-specific areas, which exacerbated the inequity in gifted identification. Gentry et al., (2019) reported that rural and town schools demonstrated considerably less equity in identification for gifted programs than city and suburban schools. They also found that nationally, students in low-SES schools were less likely to be identified as gifted students, specifically; 58% of students were from high-SES schools. Thus, after accounting for school location and school SES, examining the association between the rate of gifted program participation at school and mathematics achievement will reflect the importance of access to gifted education services.

Collectively, these studies presented thus far provide evidence that students' demographics and school contexts are associated with students' self-efficacy. Given the significance of mathematics self-efficacy in relation to mathematics achievement, it is reasonable to speculate that students' differences in mathematics self-efficacy may be associated with observed disparities in mathematics achievement among student demographic subgroups. Nonetheless, findings regarding the association of mathematics self-efficacy with achievement for different student subgroups have been variable, potentially due to limitations in sample compositions, research designs, and methodologies. Furthermore, the consistent association of mathematics self-efficacy with achievement across different grade levels has yet to be fully explored, which could be addressed with a systematic investigation with nationally representative student assessment data. As students' mathematics self-efficacy may vary with their academic or developmental progression (Schunk & Pajares, 2002), exploring the role of mathematics self-efficacy at different academic levels provides additional insight into the relationship between motivation and achievement gaps.

Thus, we examined the association between mathematics self-efficacy and mathematics achievement, taking into account student- and school-level demographic variables for three grade levels (i.e., Grades 4, 8, and 12). And we analyzed the achievement gaps by subgroups with consideration of variations in mathematics self-efficacy. Sequential two-level cross-sectional multilevel models (MLM) were applied for each grade to address the following research questions with NAEP 2019, a nationally representative large-scale dataset. The research questions are:

RQ1. What is the distribution of variability in students' mathematics achievement within schools and between schools?

RQ2. How are student-level characteristics (i.e., gender, ELL status, IEP status, NSLP eligibility, race/ethnicity, and mathematics self-efficacy) associated with student-level variations in students' mathematics achievement?

> 2.1 What proportion of the variance in mathematics achievement can be associated with mathematics self-efficacy when controlling for student-level demographic characteristics?
>
> 2.2 How do associations between student subgroups (i.e., by gender, ELL, IEP, NSLP, and race/ethnicity) and mathematics achievement differ when mathematics self-efficacy is considered?

RQ3. What are the associations between school-level characteristics and variations in student mathematics achievement?

Method

NAEP Data

This secondary data analysis study used the data from NAEP 2019 mathematics (IES license # 13090032). NAEP is designed to measure the trends in the academic performance of U.S. students in Grades 4, 8, and 12 (National Assessment Governing Board, 2020). To make valid inferences about the mathematics achievement of students in the U.S. population, NAEP employed a two-stage stratified sampling design to select students in Grades 4, 8, and 12 who were attending public, private, Bureau of Indian Education (BIE), and Department of Defense Education Activity (DoDEA) schools for assessment (National Center for Education Statistics [NCES], 2022a). The two levels of clustering in NAEP were schools within geographic groups and students within schools. Questionnaires that were administered to students, teachers, and school principals contain questions regarding students' backgrounds as well as school contexts. NAEP data are cross-sectional consisting of student- and school-level variables that are associated with students' mathematics achievement at a given point in time. The list of NAEP variable names and descriptions that are used as dependent, independent, and weights variables in this study is presented in Appendix A.

Plausible values for NAEP mathematics achievement

NAEP mathematics achievement was reported as scaled scores.[1] NAEP reported 20 plausible values of students' mathematics achievement scores (variable name in the dataset: MRPCM1-20) for estimating population characteristics from the sampled students through marginal maximum likelihood analysis (NCES, 2016). The 20 plausible values of students' mathematics achievement in NAEP 2019 served as the dependent variable as a set in this study.

Student-level predictors

Mathematics self-efficacy

Students' mathematics self-efficacy was a primary variable of interest. Several items measured students' self-efficacy in mathematics at each grade, asking students to what extent they could answer subject-specific questions. For example, there were seven items in Grade 8, one of which asked students whether they could "list all of the different possible outcomes when a coin is flipped three times." Students answered on a 5-point response scale from "I definitely cannot" to "I definitely can." Two-parameter item response theory model (NCES, 2022b) was applied to estimate scaled scores for students' self-efficacy, and the variable (SQRPM7)[2] was used for analysis. The higher the score, the higher the students' mathematics self-efficacy.

Demographic characteristics

We also used students' background variables, which included gender (DSEX), ELL status (LEP), IEP status (IEP), NSLP eligibility status (SLUNCH), and race/ethnicity (DRACE10). Gender was dummy coded. Female students were coded as 1, and male students were coded as 0. We created a set of dummy coded variables to represent the categories of race/ethnicity with White students as a reference group, that is, Asian students (Asian=1, White=0), Hispanic students (Hispanic=1, White=0), Black students (Black=1, White=0), American Indian/Alaska Native students (American Indian/Alaska Native=1, White=0), Native Hawaiian/Pacific Islander (Native Hawaiian/Pacific Islander=1, White=0), Two or more races (Two or more races=1, White=0). We collapsed the original NAEP categories of "No, formerly ELL" and "No" into the category of students who were not ELL and dummy coded it as 0 (reference group), ELLs as 1. Variables of students with IEP and 504 plans were also combined and recoded as 1, with students not having either plan as a 0. The categories of "reduced-price eligible" and "free lunch eligible" were collapsed as the group of students not eligible for NSLP (dummy coded as a 1) and students not eligible for any lunch programs as a 0.

School-level predictors

A set of background variables was selected to represent schools' characteristics. The variables of the proportion of Hispanic students (SSCHHSP), Black students (SSCHBLK), American Indian/Alaska Native students (SSCHIND), and Native Hawaiian/Pacific Islander students (SSCHHPI) are continuous, ranging from 0 to 100 percent. The values

---

[1] The scaled score ranged from 0 to 500 for Grades 4 and 8, 0 to 300 for Grade 12.

[2] NAEP reported the scaled score to represent the degree of self-efficacy.

of these four variables were added to form the new continuous variable of the proportion of underrepresented students in this study. The proportions of ELLs (C044006), students with IEPs (C044007), students eligible for NSLP (C051651), students receiving targeted Title I services[3] (C051801), and students in gifted and talented programs (C044004) are divided into several categories (for details, see Appendix A). The variable of school location (SULOCAL)[4] was recoded as a set of dummy variables with suburban schools coded as 0 (reference group) that generated three new variables, which were City (city = 1, suburb = 0), Town (town = 1, suburb = 0), and Rural (rural = 1, suburb = 0). School-level mathematics self-efficacy variable was created by taking the average of student mathematics self-efficacy in each school.

Data analysis

Because NAEP student data are nested within the school, we applied a set of two-level cross-sectional multilevel models (Raudenbush & Bryk, 2002) that enables us to examine within- and between-cluster relations of self-efficacy to mathematics achievement (Stapleton, 2013). We applied group mean centering and grand mean centering (McCoach, 2010) for independent variables to minimize bias in regression coefficient estimates (Enders & Tofighi, 2007). In this study, group means centering[5] was applied for all student-level variables. Grand mean centering[6] was used for all school-level variables. We also used the final school weight (SMSRSWT) variable at school-level models and student conditional weights (i.e., student original joint weight variable, ORIGWT/ SMSRSWT) at student-level models and to obtain unbiased students' mathematics achievement estimates representative of the target population. To assess the association between self-efficacy and the magnitude of mathematics achievement gaps, we examined the rate of variance reduction between schools captured by models (Raudenbush & Bryk, 2002). We ran all MLM analyses with HLM 8.0 software (Raudenbush & Congdon, 2021). The plausible value feature in HLM 8.0 was used to handle the 20 plausible values of mathematics achievement when running models. Maximum likelihood estimation was used for all models.

MLM Modeling steps

To address the research questions in this study, five models were tested with data from each grade level. The details about model equations for Grades 4, 8 & 12 can be found in Appendix B. First, we fitted an unconditional model (Model 0) to the data without any predictors at student or school-level, which helped to determine the extent of variance in student mathematics achievement that is associated with school-level grouping. We calculated the intraclass correlation coefficient (ICC) to assess the unit dependence

---

[3] For those public schools in which students who qualify for free or reduced lunch programs share at least 40 percent of enrollment, such schools are eligible to use federally granted Title I funds to operate schoolwide programs to enhance the achievement of students at risk (U.S. Department of Education, 2018).

[4] The dataset of grade 12 did not have this variable. It did not have any variable to explicitly represent school locale through urban-centric type, either. Therefore, in the data analysis of grade 12, there was no variables to present school locations.

[5] Group means centering refers to subtracting the mean score from the higher-level group, which is school level in this study, for all students within the same school. Thus, the between-school variations are removed from student-level predictors. The intercept presents the average score of the group students belong.

[6] Grand mean centering subtracts the overall mean for the variables across all groups from each score. The intercept represents the average score of students' mathematics achievement for all students in 2019 NAEP.

within the group. ICC here represents the proportion of total variance in student mathematics achievement accounted for by between-school heterogeneity, which ranges from 0 to 1 (Musca et al., 2011). An ICC value between 0.1 and 0.15 (Scherbaum & Ferreter, 2009) or around 0.2 for school-based clustering (Hedges & Hedberg, 2007) will support investigating the cluster effects with multilevel modeling analysis.

Next, we added the student demographic predictors into the Level-1 model (Model 1). The effects of all student-level demographic predictors were allowed to vary across schools initially but fixed for subsequent analyses when statistically non-significant random effects were detected. The students' level of self-efficacy was then incorporated into Model 1 as Model 2, to observe its association with mathematics achievement alongside demographic variables.

As the next step, we simultaneously entered the school-level demographic variables into the random slope models of Model 2 to explain the variation of the effects of the student predictors across schools (Model 3). Additionally, the school-level mean self-efficacy was included in Model 3 to observe the association of this contextual variable with the outcomes (Model 4). Finally, in Model 5, we added all the above-mentioned school-level characteristics into the Level-2 slope model for self-efficacy and tested the cross-level interactions. Note that the above analytical steps with MLM were repeated with data from Grades 4, 8, and 12, and the results were evaluated separately to address research questions. Also, note that the final conditional model was not identical across grades because some effects of student-level predictors varied across schools at one grade but not at another. In addition, since no school location information was provided in the NAEP Grade 12 dataset, we did not have the school location variables in Models 3, 4, and 5.

Findings

To minimize the repetition in reporting the findings per grade, we focus on the result of Grade 8 and highlight the key results from Grades 4 and 12. All estimates were computed based on appropriate sample size in each variable per NCES reporting guidelines of statistical results (U.S. Department of Education & Institute of Education Sciences, 2005).

Mathematics achievement variation across schools

Results with the unconditional model (Model 0) in Grade 8 showed that the grand mean of the mathematics achievement scores across schools was 281.41 ($SE = 0.57$, $p < 0.001$). However, school averages varied significantly across schools, and on average, school means deviated about 16.45 points from the overall mean ($p < 0.001$). The calculated ICC value was 0.18, which meant that about 18% of the total variance in the eighth-grade mathematics achievement scores was accounted for by between-school heterogeneity. Similar results were found in the 4th and 12th grades. For Grades 4 and 12, the grand mean of mathematics achievement (γ_{00}) is 239.52 ($SE = 0.4$, $p < 0.001$) and 148.47 ($SE = 0.67$, $p < 0.001$). But school means varied significantly between schools (Grade 4: $SD = 14.49$, variance $= 209.96$, $p < 0.001$, Grade 12: $SD = 13.38$, variance $= 178.98$, $p < 0.001$). More specifically, about 18% and 16% of the variance in NAEP 4th and 12th grade mathematics was attributable to between-school differences. While meaningful

contextual effects exist across all grade levels, the association of the contexts with student mathematics achievement appears to diminish as the grade level advances.

The effect of mathematics self-efficacy on mathematics achievement

Two-level conditional MLM models (Models 1 & 2) were applied to examine the association between mathematics self-efficacy and mathematics achievement, alongside student demographic predictors. At Grade 8, no random effect at Level-2 was statistically significant, and thus we removed all random effects from models for Level-1 coefficients at Level-2 and set the effects fixed, indicating the association of these student variables with mathematics achievement was consistent across schools. But in Grades 4 and 12, some random effects of student-level demographic predictors were detected, for example, only the intercept of IEP and NSLP demographic variables at Level-1 were allowed to vary across schools at Grade 4. Table 1 presents the results of Models 1 and 2 for Grades 4, 8, and 12. The intercept of Model 2 in Grade 8 presented that the estimated mean achievement of a student who was at the level of 0 on all dummy-coded demographic variables (i.e., non-ELL, non-NSLP While male students without IEP) was 280 points ($SE = 0.52$). A similar interpretation is applied for the conditional means for Grades 4 and 12.

The unique effect of mathematics self-efficacy after controlling demographics

The results of Model 2 showed that after controlling demographic differences, mathematics self-efficacy was significantly associated with students' mathematics achievement in Grade 8, as well as in Grades 4 and 12. Similarly, the addition of mathematics self-efficacy accounted for an extra 11.86% of the variance in Grade 4 and 17.66% in Grade 12, suggesting a substantial association between students' mathematics self-efficacy and their mathematics achievement. Furthermore, the magnitude of this association with mathematics achievement was larger in Grade 8 compared to Grades 4 and 12. An increase of one point in student-level mathematics self-efficacy was associated with an increase of 8.63 points ($SE = 0.1$) in mathematics achievement in Grade 8, 5.84 points ($SE = 0.09$) in Grade 4, and 7.34 points ($SE = 0.15$) in Grade 12. The variability in the association between mathematics self-efficacy and mathematics achievement across schools was significant for all grades, implying that this relationship was not uniform and varied by school context.

Mathematics self-efficacy and achievement gaps

Compared to the results with Model 1, the results with Model 2 showed that achievement gaps between all student subgroups (except the gendered gap at Grade 8) shrunk considerably after accounting for mathematics self-efficacy at all grades. However, the gaps were still statistically significant. Further details are described below.

Achievement gaps by gender

Gender achievement gaps in mathematics, although notable, remained relatively small in all grades in Model 1. For example, after accounting for other demographic variables, in Grade 8, the average mathematics achievement of males was 2.68 points ($SE = 0.19$, $p < 0.001$) higher than that of females. After accounting for the association

Table 1 Summary of results from the model 1 & 2 for NAEP 2019 mathematics achievement in grades 4, 8, and 12

Variables	Grade 4		Grade 8		Grade 12	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fixed effects	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
School mean mathematics achievement, γ_{00}	239** (0.36)	240.96** (0.39)	278.93** (0.49)	280** (0.52)	146.74** (0.66)	147.41** (0.69)
Student-level						
Female, γ_{10}	− 5.52** (0.24)	− 4.98** (0.25)	− 2.68** (0.19)	− 3.17** (0.28)	− 5.09** (0.48)	− 4.07** (0.45)
ELL, γ_{20}	− 15.56** (0.57)	− 12.85** (0.55)	− 28.05** (0.47)	− 21.8** (0.83)	− 29.38** (1.23)	− 25.05** (1.23)
IEP, γ_{30}	− 29.65** (0.41)	− 24.07** (0.45)	− 35.82** (0.32)	− 26.65** (0.53)	− 32.22** (0.86)	− 25.51** (0.89)
NSLP, γ_{40}	− 9.64** (0.31)	− 8.03** (0.31)	− 12.04** (0.22)	− 8.37** (0.36)	− 9.11** (0.57)	− 7.18** (0.53)
Black, γ_{50}	− 9.84** (0.43)	− 9.32** (0.44)	− 19.15** (0.37)	− 15.82** (0.57)	− 20.75** (0.8)	− 19.41** (0.8)
Hispanic, γ_{60}	− 9.07** (0.35)	− 8.22** (0.34)	− 11.45** (0.28)	− 8.02** (0.44)	− 12.71** (0.69)	− 10.17** (0.67)
Asian, γ_{70}	8.35** (0.76)	7.95** (0.78)	15.11** (0.48)	11.93** (0.8)	10.1** (1.27)	6.15** (1.13)
AIAN, γ_{80}	− 5.02** (0.89)	− 3.91** (0.87)	− 12.93** (0.99)	− 9.65** (1.35)	− 9.73** (2.64)	− 8.15** (2.5)
NHPI, γ_{90}	− 9.91** (1.39)	− 8.61** (1.4)	− 10.82** (1.46)	− 9.73** (2.16)	− 23.52** (3.45)	− 23.64** (3.5)
>1 race, γ_{100}	1.61** (0.58)	1.42** (0.59)	− 5.05** (0.42)	− 4.47** (0.61)	− 4.97** (1.12)	− 4.75** (1.05)
Self-efficacy, γ_{110}		5.84** (0.09)		8.63** (0.1)		7.34** (0.15)
Random effects	Variance	Variance	Variance	Variance	Variance	Variance
Intercept (variance between schools), u_0	240.9	240.89	319.92	339.49	227.36	252.31
IEP slope, u_3	135.05	174.23	–	–	–	–
NSLP slope, u_4	34.32	45.02	–	–	–	–
Self-efficacy slope, u_{11}	–	5.74	–	6.18	–	73.74
Student-level (variance within schools), r	527.83	346.52	860.35	567.99	695.04	495.52
Variance in achievement within schools explained (%)	29.72%	53.86%	27.65%	52.23%	26.63%	47.69%

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment

**p < .01; *p < .05

with mathematics self-efficacy, the gender gaps narrowed down slightly in Grades 4 and 12 but widened with a slight increase in Grade 8.

Achievement gaps by ELL, IEP, and NSLP

The results in Model 1 also presented a glaring mathematics achievement disparity between ELL and non-ELL students in Grades 4, 8, and 12. The disparity in Grade 4 between ELL and non-ELL students was about 15.56 points ($SE = 0.57$) when controlling other demographics. However, the disparity was widened to 28.05 points ($SE = 0.47$) in Grade 8 and 29.38 points ($SE = 1.23$) in Grade 12. Even when controlling for their mathematics self-efficacy differences, the achievement gap by ELL status at Grades 4, 8, and 12 remained relatively wide.

The original mathematics achievement gap between IEP and non-IEP students in Grade 4 was 29.65 points ($SE = 0.41$), which widened significantly with increasing school years, 35.82 points ($SE = 0.32$) in Grade 8, 32.22 points ($SE = 0.86$) in Grade 12. After considering students' mathematics self-efficacy, all gaps narrowed by about 5–9 points, but the disparities between IEP and non-IEP students were still large.

The gap between NSLP and non-NSLP students increased slightly from 9.64 points in Grade 4 ($SE = 0.31$) to 12.04 points in Grade 8 ($SE = 0.22$) and decreased to 9.11 points in Grade 12 ($SE = 0.57$). After controlling mathematics self-efficacy, the gap shrunk slightly.

Achievement gaps by race/ethnicity

The gaps of Black—White and Hispanic—White had been significant and extensive across grades, particularly for the gap of Black—White. Among Black students who had the same degree of self-efficacy in mathematics and were identical on other demographic predictors (i.e., gender, ELL, IEP, NSLP), the gaps between those Black and White students fluctuated across grades, about 9.84 points ($SE = 0.43$) in Grade 4, 19.15 points ($SE = 0.37$) in Grade 8, 20.75 points ($SE = 0.8$) in Grade 12 (see results from Model 1). After the self-efficacy difference between Black and White was controlled, the demographic effects were substantially reduced at all grades and still significant. The gaps between NHPI/AIAN and White were statistically significant regardless of their level of self-efficacy across grade levels. When considering the effects of mathematics self-efficacy, the gap between White—Asian in Grades 4, 8, and 12 remained still wide and significant, but shrunk about 4-points in Grades 8 & 12.

School contextual effects on mathematics achievement

Tables 2 and 3 summarized the results of these models for Grades 4, 8, and 12. In Model 3, the inclusion of school-level demographics explained 52.17% of the variance in average mathematic scores between schools in Grade 8, 45.9% in Grade 4, and 43.45% in Grade 12, which all suggested that in addition to students' backgrounds (including mathematics self-efficacy), the demographic composition of schools was also found to be significantly associated with school mathematics achievement. After considering all the student-level variables, all school-level demographic variables, except the percentage of ELLs and city schools or not in Grade 8, significantly explain the variation in school average mathematics achievement. Taking Grade 8 as an example, as for school location, in comparison with suburban schools with similar school settings (i.e., percentage of ELL, IEP, students receiving Title I service, gifted students, and underrepresented students), school average mathematic achievement

was 9.95 points ($SE = 1.14$) lower for town schools; 8.99 points ($SE = 0.94$) lower for rural schools. Also, for example, a one-unit increase (e.g., from 26–50% to 51–75%) in the percentage of IEP students in schools would result in a 3.82-point ($SE = 0.52$) decrease in school mean achievement with all other contextual backgrounds equal. A one-point increase in the percentage of underrepresented students in schools also resulted in 0.33 points ($SE = 0.02$) decrease in school mean achievement. All differences are statistically significant.

The results with Model 4 indicated that school mean mathematics self-efficacy was significantly associated with the school-level mean mathematics achievement after controlling school-level demographics across grades. This finding indicates a potential contextual correlation with self-efficacy, as the association of the average school mathematics self-efficacy with achievement was observed to be more pronounced than that of individual self-efficacy variables across the grade levels. For Grade 8, a one-point change in the average school-level mathematics self-efficacy is associated with a 17.4-point difference ($SE = 0.42$) in the average school mathematics achievement; for Grade 4, there is a 14.46-point difference ($SE = 0.51$); and for Grade 12, a 14-point difference ($SE = 0.76$). Including the average school mathematics self-efficacy in Model 4 accounted for an additional 27.56% of the between-school variance in Grade 8, 20.42% in Grade 4, and 27.42% in Grade 12, suggesting a substantial correlation. Although we were originally interested in whether school-level characteristics, specifically school average self-efficacy, explain the variation in the relationship between mathematics self-efficacy and mathematics achievement across schools, we did not find significant variation in the relationship in Model 2. Accordingly, as reported in Table 3, the results showed trivial changes in variances between schools explained by Model 5 compared to Model 4 across three grades.

Discussion

Mathematics achievement disparities by students' socio-demographic subgroups have been widely discussed for decades (Hanushek et al., 2019), which instead of contributing to shrinking the achievement gaps but reinforced the negative stereotypes thrown on traditionally underrepresented students (Garcia & Guerra, 2004). In this study, using asset-oriented perspectives on reducing achievement gaps by subgroups, we draw attention away to the role of malleable characteristics of students (i.e., mathematics self-efficacy), which is a significant motivational process and predicts subsequent academic outcomes (Schunk & DiBenedetto, 2021). Our aim was to determine how mathematics self-efficacy is associated with mathematics achievement across different developmental stages and to consider its potential role as a factor related to the achievement gap attributable to students' backgrounds. The findings indicate that mathematics self-efficacy is correlated with achievement, with the association appearing most prominent in Grade 8 compared to Grades 4 and 12. Contextual factors such as the proportion of students eligible for NSLP and from underrepresented groups, as well as school location, are also shown to be related to achievement, resonating with previous research (Pitsia et al., 2017; Wu, 2015). The findings highlight the relevance of learning; or school contexts when understanding students' achievement. Additionally, the data suggest that mathematics self-efficacy is related to the narrowing of achievement disparities across

Table 2 Summary of results from the model 3 & 4 for NAEP 2019 mathematics achievement in grades 4, 8, and 12

Variables	Grade 4		Grade 8		Grade 12	
	Model 3	Mode 4	Model 3	Mode 4	Model 3	Mode 4
Fixed effects	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
School mean mathematics achievement, γ_{00}	240.99** (0.29)	231.09** (0.4)	281.48** (0.41)	270.21** (0.41)	147.48** (0.57)	146.77** (0.45)
Student-level						
Female, γ_{10}	− 5** (0.26)	− 5.01** (0.26)	− 3.24** (0.28)	− 3.25** (0.28)	− 4.03** (0.46)	− 4.09** (0.46)
ELL, γ_{20}	− 12.75** (0.57)	− 12.73** (0.57)	− 21.58** (0.83)	− 21.58** (0.83)	− 24.67** (1.23)	− 24.62** (1.23)
IEP, γ_{30}	− 24.19** (0.46)	− 24.18** (0.646)	− 26.61** (0.54)	− 26.62** (0.55)	− 25.52** (0.92)	− 25.47** (0.92)
NSLP, γ_{40}	− 8.01** (0.31)	− 8.01** (0.31)	− 8.28** (0.36)	− 8.28** (0.36)	− 7.35** (0.54)	− 7.33** (0.54)
Black, γ_{50}	− 9.49** (0.44)	− 9.748** (0.44)	− 15.92** (0.57)	− 15.91** (0.57)	− 19.51** (0.83)	− 19.48** (0.83)
Hispanic, γ_{60}	− 8.19** (0.35)	− 8.19** (0.35)	− 8.17** (0.45)	− 8.17** (0.45)	− 10.01** (0.68)	− 10.01** (0.68)
Asian, γ_{70}	8.01** (0.78)	8.02** (0.78)	11.91** (0.83)	11.98** (0.83)	6.97** (1.17)	6.93** (1.17)
AIAN, γ_{80}	− 4.04** (0.89)	− 4.03** (0.89)	− 9.54** (1.36)	− 9.47** (1.36)	− 7.71** (2.58)	− 7.67** (2.58)
NHPI, γ_{90}	− 8.81** (1.45)	− 8.8** (1.44)	− 9.9** (2.21)	− 9.86** (2.21)	− 24.19** (3.57)	− 24.17** (3.54)
>1 race, γ_{100}	1.47** (0.61)	1.48** (0.61)	− 4.68** (0.63)	− 4.66** (0.63)	− 4.77** (1.09)	− 4.74** (1.09)
Self-effi-cacy, γ_{110}	5.85** (0.09)	5.86** (0.09)	8.75** (0.1)	8.75** (0.1)	7.33** (0.15)	7.37** (0.15)
School-level for β_0						
% of ELL, γ_{01}	0.55* (0.24)	0.81** (0.19)	0.54 (0.37)	1.3** (0.28)	0.97* (0.47)	0.52 (0.45)
% of IEP, γ_{02}	− 2.49** (0.38)	− 1.76** (0.3)	− 3.82** (0.52)	− 2.06** (0.37)	− 5.18** (0.75)	− 3.42** (0.58)
% of Title I, γ_{03}	− 1.01** (0.11)	− 0.76** (0.1)	− 1.325** (0.15)	− 0.62** (0.11)	− 0.91** (0.23)	− 0.6** (0.19)
% of gifted students, γ_{04}	2.91** (0.25)	1.8** (0.21)	2.35** (0.28)	1.01** (0.19)	1.62** (0.42)	1.06** (0.32)
% of underrep-resented students, γ_{05}	− 0.26** (0.01)	− 0.2** (0.01)	− 0.33** (0.02)	− 0.22** (0.01)	− 0.27** (0.02)	− 0.22** (0.02)
City, γ_{06}	− 1.86* (0.72)	− 1.56** (0.58)	− 1.16(1.07)	− 2.33** (0.7)		
Town, γ_{07}	− 5.99** (0.99)	− 4.8** (0.77)	− 9.95** (1.14)	− 5.26** (0.81)		
Rural, γ_{08}	− 6.14** (0.7)	− 5.03** (0.58)	− 8.99** (0.94)	− 3.76** (0.7)		
Mean self-efficacy, γ_{09}		14.46** (0.51)		17.4** (0.42)		14** (0.76)
Random Effects	Variance	Variance	Variance	Variance	Variance	Variance
Intercept (variance between schools), u_0	130.32	81.12	162.39	68.8	142.68	73.51

Table 2 (continued)

Variables	Grade 4		Grade 8		Grade 12	
	Model 3	Mode 4	Model 3	Mode 4	Model 3	Mode 4
Fixed effects	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
IEP slope, u_3	190.56	190.47	–	–	–	–
NSLP slope, u_4	51.41	51.5	–	–	–	–
Self-efficacy slope, u_{11}	6.31	6.33	6.42	6.37	3.7	3.84
Student-level (variance within schools), r	323.46	323.62	523.67	523.61	460.9	462.07
Variance in achievement between schools explained (%)	45.9%	66.32%	52.17%	79.73%	43.45%	70.87%

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment

$**p < .01$; $*p < .05$

socio-demographic subgroups. In the subsequent sections, we will delve into these associations in greater detail and discuss the implications of our findings for educational practice and future research, while acknowledging that our cross-sectional study design precludes the establishment of causal relationships.

The association between mathematics self-efficacy and achievement

Consistent with previous research (e.g., Bohrnstedt et al., 2020; Keşan & Kaya, 2018), findings from this study highlight the significant correlation between students' mathematics self-efficacy and mathematics achievement across grades. Notably, results indicate that mathematics self-efficacy had the largest association with achievement in Grade 8. This observation points to the potential importance of supporting middle school students in bolstering their mathematics self-efficacy, which may be related to their current and prospective mathematics performance. We posit that there could be at least two reasons for this observed correlation.

First, as students progress through their academic coursework into middle school, they become more aware of the bombarded feedback cues about their mathematics performance that are from their primary resources (e.g., peers, mathematics teachers, and parents) (Hickman & Sherman, 2019). With the increasing complexity of mathematics in middle school, the positive cues students received from their direct ecological environments that blend with their positive dispositions toward learning mathematics play a determinant role in forming a high level of self-efficacy in mathematics (Hickman & Sherman, 2019). For example, for students who have a history of outstanding mathematics performance and access to mathematics learning facilities/resources, they tend to feel mathematics is becoming manageable in middle school and self-affirm their

Table 3 Summary of results from the model 5 for NAEP 2019 mathematics achievement in grades 4, 8, and 12

Variables Fixed effects	Grade 4 Model 5: Full Model Estimate (*SE*)	Grade 8 Model 5: Full Model Estimate (*SE*)	Grade 12 Model 5: Full Model Estimate (*SE*)
School mean mathematics achievement, γ_{00}	231.15** (0.4)	270.12** (0.41)	146.66** (0.45)
Student-level			
Female, γ_{10}	− 4.97** (0.26)	− 3.23** (0.28)	− 4.05** (0.46)
ELL, γ_{20}	− 12.75** (0.57)	− 21.71** (0.83)	− 24.72** (1.24)
IEP, γ_{30}	− 24.12** (0.46)	− 26.58** (0.55)	− 25.36** (0.92)
NSLP, γ_{40}	− 7.96** (0.31)	− 8.22** (0.36)	− 7.27** (0.54)
Black, γ_{50}	− 9.42** (0.44)	− 15.87** (0.57)	− 19.41** (0.83)
Hispanic, γ_{60}	− 8.16** (0.35)	− 8.11** (0.45)	− 10.02** (0.68)
Asian, γ_{70}	8** (0.78)	11.86** (0.83)	6.9** (1.17)
AIAN, γ_{80}	− 4.02** (0.89)	− 9.56** (1.36)	− 7.63** (2.59)
NHPI, γ_{90}	− 8.75** (1.44)	− 9.92** (2.22)	− 23.94** (3.52)
> 1 race, γ_{100}	1.46* (0.61)	− 4.64** (0.63)	− 4.74** (1.09)
Self-efficacy, γ_{110}	6.15** (0.17)	8.25** (0.17)	7.25** (0.16)
% of ELL, γ_{111}	0.2* (0.08)	0.31** (0.1)	0.42** (0.14)
% of IEP, γ_{112}	− 0.01 (0.112)	− 0.07 (0.15)	− 0.1 (0.2)
% of Title I, γ_{113}	− 0.08* (0.04)	− 0.09* (0.04)	− 0.08 (0.07)
% of gifted students, γ_{114}	0.05 (0.07)	0.01 (0.07)	0.05 (0.12)
% of underrepresented students, γ_{115}	− 0.02** (0.01)	− 0.02** (0.01)	− 0.03** (0.01)
City, γ_{116}	0.33 (0.24)	− 0.1 (0.28)	
Town, γ_{117}	− 0.07 (0.29)	− 0.07 (0.3)	
Rural, γ_{118}	− 0.44 (0.25)	− 0.55* (0.26)	
Mean self-efficacy, γ_{119}	− 0.44* (0.2)	0.64** (0.17)	− 0.26 (0.27)
School-level for β_0			
% of ELL, γ_{01}	0.86** (0.2)	1.36** (0.28)	0.82 (0.42)
% of IEP, γ_{02}	− 1.75** (0.3)	− 2.06** (0.37)	− 3.47** (0.62)
% of Title I, γ_{03}	− 0.78** (0.1)	− 0.64** (0.11)	− 0.67** (0.19)
% of gifted students, γ_{04}	1.81** (0.2)	1.02** (0.19)	1.09** (0.33)
% of underrepresented students, γ_{05}	− 0.2** (0.01)	− 0.23** (0.01)	− 0.24** (0.02)
City, γ_{06}	− 1.48* (0.59)	− 2.36** (0.71)	–
Town, γ_{07}	− 4.82** (0.78)	− 5.3** (0.81)	–
Rural, γ_{08}	− 5.15** (0.59)	-3.9** (0.7)	–
Mean self-efficacy, γ_{07}	14.36** (0.51)	17.54** (0.42)	13.89** (0.78)
Random Effects	Variance	Variance	Variance
Intercept (variance between schools), u_0	81.09	68.81	73.2
IEP slope, u_3	189.88	–	–
NSLP slope, u_4	51.9	–	–
Self-efficacy slope, u_{11}	5.85	5.7	3.43
Student-level (variance within schools), r	323.52	523.43	461.82
Variance in achievement between schools explained (%)	66.34%	79.73%	70.99%

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment

**$p < .01$; *$p < .05$

mathematics abilities (Usher, 2009). It is posited that the level of mathematics self-efficacy is correlated with students' self-regulation, persistence, engagement, and effort in tackling challenging mathematics problems, which in turn are associated with differences in mathematics achievement (Bandura, 1977).

Second, when students go to middle school, students' mathematics experiences start to diverge because of tracking and course placement decisions. Based on Subotnik et al. (2021)'s talent development megamodel, when students enter early adolescence, their mathematics talent begins to flourish with the provision of advanced mathematics learning opportunities and psychosocial skills (e.g., self-efficacy) (Subotnik et al., 2021). Middle school opens the door for able students to embrace multiple spheres of opportunities (e.g., honors classes, math clubs, academic summer camps) from which they build high self-efficacy through discourses and interactions with others. Unfortunately, for students not placed in advanced mathematics classes, they often face the prospects of less rigorous coursework, lack of teacher support, and low expectation, whereas their able peers are taught coherent conceptual understanding and high-order thinking skills (Stiff et al., 2011). Additionally, research has shown a correlation between placement in lower academic tracks and a decrease in self-efficacy, which is associated with changes in cognition and performance (Gray et al., 2002). Furthermore, this situation appears to be correlated with disproportionate effects on underrepresented students (e.g., Black, Hispanic, Native American, students from low-income families, ELLs, and students with disabilities), with some indicators that placement decisions may be associated with demographic data (Stiff et al., 2011).

Mathematics self-efficacy and achievement gaps

When students have similar levels of self-efficacy in mathematics, we observed that the associations of demographic factors with their achievement were less pronounced; mathematics achievement gaps by subgroups were reduced from small to large extent. This implies that mathematics self-efficacy could be uniformly associated with supporting mathematics achievement, irrespective of subgroup identity. But some student subgroups tend to have low mathematics self-efficacy, which is associated with lower achievement outcomes.

In our study, the correlation between mathematics self-efficacy and the narrowing of gender and racial-ethnic achievement gaps in mathematics was not significant, whereas a strong positive correlation with self-efficacy was observed in the achievement gaps of ELL, IEP, and NSLP students, particularly in Grade 8, which is in line with the findings of Polat et al. (2016) and Soland and Sandilos (2021). With the statistical control of differences in mathematics self-efficacy, we noted a correlation with the reduction of achievement gaps, leading to the consideration that the gaps often seen between well-represented and underrepresented students are associated with differences in self-efficacy levels. According to Bandura's (1977) sources of self-efficacy, ELLs, students with IEPs, and students eligible for NSLP are traditionally documented to experience persistent underperformance in mathematics (Rodriguez et al., 2022) (mastery experiences), express frustrations after social comparisons to normally functioning peers (Coleman, 2001) (vicarious experiences), receive few encouragements (Solomon et al., 1996) (verbal persuasion), and struggle with physical and psychological ill-being (Campbell & Gilmore, 2014) (emotional and physiological states). These factors might interact with the

context of having under-proficient mathematics teachers (Abedi et al., 2006), who may not effectively reinforce concepts of mathematics self-efficacy, or being in environments with prevalent negative stereotypes (Steele & Aronson, 1995), or attending schools in high-poverty areas with limited resources (Darling-Hammond, 2013), which are all associated with lower self-efficacy appraisals.

In contrast, White and Asian students, non-ELLs, students without learning disabilities, and students from families with more resources are often associated with a wealth of reinforcing experiences for self-efficacy at schools and homes, such as skillful mathematics teachers structured routine opportunities for success in mathematics and provide above-grade-level mathematics instructions (Usher, 2009); these experiences are correlated with higher self-efficacy in mathematics. Consequently, if ELL, IEP, and NSLP students were to exhibit high levels of self-efficacy comparable to their peers, they too might show similar academic achievements, and the observed achievement gaps might be less pronounced. It indicates that exploring how underrepresented groups cultivate an interest in mathematics, form aspirations related to mathematics, construct effective responses to stereotypes, and see themselves as capable mathematicians, be correlated with enhancements in their self-efficacy in mathematics.

Contextual correlations with demographic factors

The findings from this study also indicate a significant association between environmental factors (e.g., proportions of students eligible for NSLP and racially/ethnically underrepresented students, the proportion of gifted students, and school location) and the overall mathematics achievement in school. With the findings indicating that schools with a high proportion of students from low-income families, as well as Black, Hispanic, and Native American students, tend to have lower mathematics achievement compared to others. This underscores the importance of exploring how certain schools with comparable demographic profiles are associated with higher levels of student success in mathematics despite facing academic, social, and institutional challenges. For example, in-depth qualitative research into the student support strategies employed in such schools could yield insights that are correlated with the enhancement of educational outcomes for students from diverse cultural and racial backgrounds.

This underscores the importance of further research into scaling up interventions and practices that are associated with effectiveness in various locales. For example, findings from this study are associated with the idea that increased access to gifted education services correlates with smaller mathematics achievement gaps. It aligns with the finding by Young et al. (2017) with NAEP (2009) data that there is no significant difference in mathematics achievement between Black and White fourth-grade girls when Black students had similar levels of access to participate in gifted and talented programs.

Implications for practices

Previous studies (e.g., Plucker et al., 2013; Yang & Maeda, 2023) have identified the emergence of the mathematics achievement gap as early as fourth grade. This study finds a significant association between mathematics self-efficacy and mathematics achievement beginning from fourth grade. This highlights the correlation between early mathematics self-efficacy and achievement outcomes and suggests the potential benefits of supporting

students' mathematics self-efficacy during early school years in relation to achievement gaps. The change in students' mathematics self-efficacy is associated with both internal personal and external environmental conditions, which can be variable and controllable (Van der Biji & Shortridge-Baggett, 2001). Likewise, students' self-efficacy development in mathematics is associated with school contexts and involves interactions between students and educational elements such as effective mathematics teaching, high-quality mathematics programs, supportive school environments, and stakeholders (e.g., school leaders, mathematics teachers, and specialists) (e.g., Bobis et al., 2013; Griggs et al., 2013; Johnsen & Sheffield, 2021). Thus, fostering students' resilient sense of self-efficacy in mathematics is associated with continued support from schools and educators at all levels, from elementary to high schools.

As the correlations observed in our study suggest, there is a potential benefit in focusing efforts toward rehumanizing mathematics education, particularly for students who are traditionally underrepresented in mathematics (Goffney et al., 2018). For example, it may be associated with positive outcomes if public schools frontload high-quality mathematics education since early childhood education (Plucker et al., 2017) to build young students' self-efficacy in mathematics. Teachers create inclusive and culturally responsive mathematics learning environments that involve underrepresented students with high potential in meaningful and rigorous learning activities (Yang & Gentry, 2023), which are correlated with increased mastery experiences and thus may support self-efficacy. In this regard, we suggest revisiting the approach of anti-deficit achievement theory (Harper, 2010) and concentrating our conversations on closing achievement gaps by understanding how students from diverse cultural, linguistic, and socioeconomic backgrounds, are associated with high levels of self-efficacy and how this relates to navigating the systematic environmental barriers put on them. This would be more meaningful than exhaustive and overt statements of achievement gaps among student subgroups.

Limitations and future research directions

The large sample size and representative nature of the final analytic sample drawn from the NAEP dataset support the validity of the results reported in this study. However, some possible limitations are unavoidable, which may require some caution for implications and may be addressed in future studies.

First, the NAEP dataset used in this study is cross-sectional; therefore, we refrain from inferring any long-term influences of mathematics self-efficacy on mathematics achievement over time. Consequently, while our findings reveal various patterns in the association between self-efficacy and achievement, we advise researchers to employ longitudinal large-scale assessment data to more accurately discern the trends of demographic characteristics and self-efficacy in relation to mathematics achievement throughout the course of education. Additionally, qualitative research could provide valuable insights into the sources of self-efficacy among underrepresented students in mathematics.

Second, mathematics self-efficacy cannot capture all variations in mathematics achievement disparities within schools. Further research could examine the correlations between the mathematics achievement gaps among student subgroups and additional student motivation-related variables (e.g., achievement goal orientations, interest, enjoyment in mathematics, persistence, and effort expended). Future studies should also be

mindful of the association of collective self-efficacy[7] at the school level. There is a trend that students are engaged in cooperative learning mathematics projects and teacher designed mathematics curricula through collaborative work (Schunk & DiBenedetto, 2021). Therefore, examining how collective self-efficacy is related to students' mathematics achievement could shed light on promoting students' self-efficacy and mathematics achievement. In addition, research has indicated that the teachers' collective self-efficacy[8] in teaching is significantly associated with academic performance and the educational environment of students, even after controlling for previous student accomplishments and crucial demographic variables, such as socioeconomic status (e.g., Bandura, 1993). Schools that achieve success are identified by teachers' shared beliefs in the capabilities of their staff to assist students in their growth and learning (Klassen et al., 2011). Therefore, investigating the factors contributing to mathematics teachers' collective self-efficacy, especially in schools with scarce resources and a high proportion of underrepresented students, and how this is associated with their students' mathematics abilities, could contribute valuable insights.

Finally, mathematics performance disparities are associated with more than just the demographic composition of schools. Findings from other NAEP studies have shown that contextual factors, such as teachers' professional development (Havard et al., 2018), classroom instruction (Lubienski, 2006), and teacher resources (Lee & Reeves, 2012), are also associated with students' self-efficacy and mathematics achievement. Future researchers may consider including a broader spectrum of school-level variables to understand these contextual correlations further. Moreover, we also encourage future researchers to include some equity-related variables in mathematic learning as part of contextual variables. In this study, we considered the variable representing students' access to gifted programs, acknowledging that such variables need further careful operationalization and modeling to enhance their interpretative validity.

Conclusion

The purpose of this study was to explore the association between mathematics self-efficacy and mathematics achievement disparities. As an exploratory study, it sheds light on how enhancing mathematics self-efficacy is related to the achievement gaps among student subgroups. We found a significant correlation between mathematics self-efficacy and U.S. students' mathematics achievement, particularly noting a larger association in Grade 8. It provides empirical evidence for supporting the asset-oriented approach to consider students' mathematics self-efficacy as a potentially influential and malleable factor in relation to achievement gaps.

---

[7] Bandura (1977) defined collective efficacy as "a group's shared belief in its conjoint capabilities to organize and execute the courses of action required to produce given levels of attainments" (p. 477).

[8] Teachers' collective efficacy pertains to their conviction in their collaborative capacity to affect the lives of their students (Bandura, 1993).

Appendix 1

Dependent, independent, and weight variables in MLM

Construct	NAEP 2019 data variable name	Description of the variable	Coding/score range
Mathematics Achievement	MRPCM1-20	20 mathematics composite plausible values	Gr. 4: 0–500 Gr. 8:0–500 Gr.12: 0–300
Student level			
Gender	DSEX	Gender	1 = Female, 0 = Male
Race/ethnicity	DRACE10	Race/ethnicity (student-reported)	1 = Asian, not Hispanic 0 = White; 1 = Hispanic of any race, 0 = White; 1 = Black, not Hispanic, 0 = White; 1 = American Indian/Alaska Native (AIAN), 0 = White; 1 = Native Hawaiian/ Pacific Islander (NHPI), 0 = White; 1 = Two or more races, 0 = White;
ELL	ELL	Student classified as English language learner	1 = ELL, 0 = Not ELL
IEP	IEP	Student classified as having a disability	1 = students with IEPs, 0 = students without IEPs
NSLP	SLUNCH	National School Lunch Program eligibility	1 = Yes (eligible for NSLP program) 0 = No (not eligible for NSLP program)
Mathematics self-efficacy	SQRPM7	Students' self-efficacy in mathematics	Gr.4:6.62—12.96 Gr.8:5.63—13.37 Gr.12: 5.54–14.14
School contextual level			
Proportion_ ELLs	C044006	Percent receiving ESL instruction	0 = None, 1 = 1–5%, 2 = 6–10%, 3 = 11–25%, 4 = 26–50%, 5 = 51–75%, 6 = 76–90%, 7 = Over 90%, 88 = Omitted
Proportion_ IEPs	C044007	Percent in special education	0 = None, 1 = 1–5%, 2 = 6–10%, 3 = 11–25%, 4 = 26–50%, 5 = 51–75%, 6 = 76–90%, 7 = Over 90%, 88 = Omitted
Proportion_Title I service	C051801	Percent receiving targeted Title I services	0 = None, 1 = 1–5%, 2 = 6–10%, 3 = 11–25%, 4 = 26–50%, 5 = 51–75%, 6 = 76–90%, 7 = Over 90%, 88 = Omitted
Proportion_gifted	C044004	Percent in gifted and talented program	0 = None, 1 = 1–5%, 2 = 6–10%, 3 = 11–25%, 4 = 26–50%, 5 = 51–75%, 6 = 76–90%, 7 = Over 90%, 88 = Omitted
Proportion_Hispanic	SSCHHSP	School percent Hispanic	0–100%
Proportion_ Black	SSCHBLK	School percent Black	0–100%
Proportion_ AIAN	SSCHIND	School percent American Indian/Alaska Native	0–100%

Construct	NAEP 2019 data variable name	Description of the variable	Coding/score range
Proportion_ NHPI	SSCHHPI	School percent Native Hawaiian/ Pacific Islander	0–100%
Proportion_ NSLP	C051651	Percent eligible National School Lunch Program	0 = 0%, 2 = 1–5%,3 = 6–10%,4 = 11–25%,5 = 26–34%,6 = 35–50%,7 = 51–75%,8 = 76–99%, 9 = 100%, 88 = Omitted
School-location	SULOCAL	Urban-centric type of locale	1 = City, 0 = Suburb; 1 = Town, 0 = Suburb; 1 = Rural, 0 = Sub-urb
Mean mathematics self-efficacy*	–	Mean of students' mathematics self-efficacy	–
Weights			
	ORIGWT	Student original joint weight	1.8751–553.7003
	SMSRSWT	School final weight	1–855.7789

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Grade 4, 8, and 12 Mathematics Assessment

Variable with * was calculated in this study

Appendix 2

Model equations across grades 4, 8 &12

Model 0 equation: unconditional model

	Grade 4	Grade 8	Grade 12
Level-1 Model	MRPCM1 = B0 + r		
Level-2 Model	B0 = G00 + u0		

Model 1 Equation: add all student-level demographic variables

	Grade 4	Grade 8	Grade 12
Level-1 Mode	MRPCM1 = B0 + B1* (FEMALE) + B2* (ELL) + B3* (IEP) + B4* (NSLP) + B5* (BLACK) + B6* (HIS-PANIC) + B7* (ASIAN) + B8* (AIAN) + B9* (NHPI) + B10* (> 1 RACE) + r		
Level-2 Model	B0 = G00 + u0	B0 = G00 + u0	B0 = G00 + u0
	B1 = G10	B1 = G10	B1 = G10
	B2 = G20	B2 = G20	B2 = G20
	B3 = G30 + u3	B3 = G30	B3 = G30
	B4 = G40 + u4	B4 = G40	B4 = G40
	B5 = G50	B5 = G50	B5 = G50
	B6 = G60	B6 = G60	B6 = G60
	B7 = G70	B7 = G70	B7 = G70
	B8 = G80	B8 = G80	B8 = G80
	B9 = G90	B9 = G90	B9 = G90
	B10 = G100	B10 = G100	B10 = G100

Model 2 Equation: add self-efficacy variable

	Grade 4		Grade 8	Grade 12
Level-1 Model	MRPCM1 = B0 + B1* (FEMALE) + B2* (ELL) + B3* (IEP) + B4* (NSLP) + B5* (BLACK) + B6* (HISPANIC) + B7* (ASIAN) + B8* (AIAN) + B9* (NHPI) + B10* (> 1 RACE) + B11* (SELF-EFFICACY) + r			
Level-2 Model	B0 = G00 + u0		B0 = G00 + u0	B0 = G00 + u0
	B1 = G10		B1 = G10	B1 = G10
	B2 = G20		B2 = G20	B2 = G20
	B3 = G30 + u3		B3 = G30	B3 = G30
	B4 = G40 + u4		B4 = G40	B4 = G40
	B5 = G50		B5 = G50	B5 = G50
	B6 = G60		B6 = G60	B6 = G60
	B7 = G70		B7 = G70	B7 = G70
	B8 = G80		B8 = G80	B8 = G80
	B9 = G90		B9 = G90	B9 = G90
	B10 = G100		B10 = G100	B10 = G100
	B11 = G110 + u11		B11 = G110 + u11	B11 = G110 + u11

Model 3 Equation: add all school-level demographic variables to the random slope

	Grade 4	Grade 8	Grade 12
Level-1 Model	MRPCM1 = B0 + B1* (FEMALE) + B2* (ELL) + B3* (IEP) + B4* (NSLP) + B5* (BLACK) + B6* (HISPANIC) + B7* (ASIAN) + B8* (AIAN) + B9* (NHPI) + B10* (> 1 RACE) + B11* (SELF-EFFICACY) + r		
Level-2 Model	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)
	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL) + u0	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL) + u0	+ G05* (% UNDERREPRESENTED) + u0
	B1 = G10	B1 = G10	B1 = G10
	B2 = G20	B2 = G20	B2 = G20
	B3 = G30 + u3	B3 = G30	B3 = G30
	B4 = G40 + u4	B4 = G40	B4 = G40
	B5 = G50	B5 = G50	B5 = G50
	B6 = G60	B6 = G60	B6 = G60
	B7 = G70	B7 = G70	B7 = G70
	B8 = G80	B8 = G80	B8 = G80
	B9 = G90	B9 = G90	B9 = G90
	B10 = G100	B10 = G100	B10 = G100
	B11 = G110 + u11	B11 = G110 + u11	B11 = G110 + u11

Model 4 Equation: add school-level mean self-efficacy variable to the random slope

	Grade 4	Grade 8	Grade 12
Level-1 Model	MRPCM1 = B0 + B1* (FEMALE) + B2* (ELL) + B3* (IEP) + B4* (NSLP) + B5* (BLACK) + B6* (HISPANIC) + B7* (ASIAN) + B8* (AIAN) + B9* (NHPI) + B10* (> 1 RACE) + B11* (SELF-EFFICACY) + r		
Level-2 Model	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)
	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL)	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL)	+ G05* (% UNDERREPRESENTED)
	+ G09* (SELF-EFFICACY_M) + u0	+ G09* (SELF-EFFICACY_M) + u0	+ G06* (SELF-EFFICACY_M) + u0
	B1 = G10	B1 = G10	B1 = G10
	B2 = G20	B2 = G20	B2 = G20
	B3 = G30 + u3	B3 = G30	B3 = G30
	B4 = G40 + u4	B4 = G40	B4 = G40
	B5 = G50	B5 = G50	B5 = G50
	B6 = G60	B6 = G60	B6 = G60
	B7 = G70	B7 = G70	B7 = G70
	B8 = G80	B8 = G80	B8 = G80
	B9 = G90	B9 = G90	B9 = G90
	B10 = G100	B10 = G100	B10 = G100
	B11 = G110 + u11	B11 = G110 + u11	B11 = G110 + u11

Model 5 Equation: add all school-level variables into level-2 self-efficacy slope model

	Grade 4	Grade 8	Grade 12
Level-1 Model	MRPCM1 = B0 + B1* (FEMALE) + B2* (ELL) + B3* (IEP) + B4* (NSLP) + B5* (BLACK) + B6* (HISPANIC) + B7* (ASIAN) + B8* (AIAN) + B9* (NHPI) + B10* (> 1 RACE) + B11* (SELF-EFFICACY) + r		
Level-2 Model	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)	B0 = G00 + G01* (%ELL) + G02* (%IEP) + G03* (% TITLE I) + G04* (% GIFTED)
	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL)	+ G05* (% UNDERREPRESENTED) + G06* (CITY) + G07* (TOWN) + G08* (RURAL)	+ G05* (% UNDERREPRESENTED)
	+ G09* (SELF-EFFICACY_M) + u0	+ G09* (SELF-EFFICACY_M) + u0	+ G06* (SELF-EFFICACY_M) + u0
	B1 = G10	B1 = G10	B1 = G10
	B2 = G20	B2 = G20	B2 = G20
	B3 = G30 + u3	B3 = G30	B3 = G30
	B4 = G40 + u4	B4 = G40	B4 = G40
	B5 = G50	B5 = G50	B5 = G50
	B6 = G60	B6 = G60	B6 = G60
	B7 = G70	B7 = G70	B7 = G70
	B8 = G80	B8 = G80	B8 = G80
	B9 = G90	B9 = G90	B9 = G90
	B10 = G100	B10 = G100	B10 = G100
	B11 = G110 + G111* (%ELL) + G112* (%IEP) + G113* (% TITLE I) + G114* (% GIFTED)	B11 = G110 + G111* (%ELL) + G112* (%IEP) + G113* (% TITLE I) + G114* (% GIFTED)	B11 = G110 + G111* (%ELL) + G112* (%IEP) + G113* (% TITLE I) + G114* (% GIFTED)
	+ G115* (% UNDERREPRESENTED) + G116* (CITY) + G117* (TOWN) + G118* (RURAL)	+ G115* (% UNDERREPRESENTED) + G116* (CITY) + G117* (TOWN) + G118* (RURAL)	+ G115* (% UNDERREPRESENTED)
	+ G119* (SELF-EFFICACY_M) + u11	+ G119* (SELF-EFFICACY_M) + u11	+ G116* (SELF-EFFICACY_M) + u11

Declarations

References

Abedi, J., Courtney, M., Leon, S., Kao, J., Azzam, T. (2006). English language learners and math achievement: A study of opportunity to learn and language accommodation. Technical report 702. National center for research on evaluation, standards, and student testing.

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability students: a comparison between Black, Hispanic, and White students. *Science Education, 98*(2), 216–242. https://doi.org/10.1002/sce.21092

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117–148.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bobis, J., Mulligan, J., & Lowrie, T. (2013). *Mathematics for children: Challenging children to think mathematically*. London: Pearson.

Bohrnstedt, G. W., Zhang, J., Park, B. J., Ikoma, S., Broer, M., & Ogut, B. (2020). Mathematics identity, self-efficacy, and interest and their relationships to mathematics achievement: A longitudinal analysis. In R. T. Serpe, R. Stryker, & B. Powell (Eds.), *Identity and symbolic interaction: Deepening foundations, building bridges* (pp. 169–210). Springer.

Campbell, M., & Gilmore, L. (2014). The importance of social support for students with intellectual disability: An intervention to promote mental health and well-being. *Cypriot Journal of Educational Sciences, 9*(1), 21–28.

Cheema, J. R., & Galluzzo, G. (2013). Analyzing the gender gap in math achievement: Evidence from a large-scale US sample. *Research in Education, 90*(1), 98–112.

Cheema, J. R., & Kitsantas, A. (2014). Influences of disciplinary classroom climate on high school student self-efficacy and mathematics achievement: A look at gender and racial–ethnic differences. *International Journal of Science and Mathematics Education, 12*(5), 1261–1279. https://doi.org/10.1007/s10763-013-9454-4

Coleman, M. R. (2001). Surviving or thriving? 21 gifted boys with learning disabilities share their school stories. *Gifted Child Today, 24*(3), 56–64. https://doi.org/10.4219/gct-2001-538

Dai, D. Y., & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly, 57*(3), 151–168. https://doi.org/10.1177/0016986213490020

Darling-Hammond, L. (2013). Inequality and school resources. In K. G. Welner & P. L. Carter (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 77–93). Oxford University Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Garcia, S. B., & Guerra, P. L. (2004). Deconstructing deficit thinking: Working with educators to create more equitable learning environments. *Education and Urban Society, 36*(2), 150–168.

Gentry, M., Gray, A., Whiting, G. W., Maeda, Y., & Pereira, N. (2019). Access denied/system failure: Gifted education in the United States: Laws, access, equity, and missingness across the country by locale, Title I school status, and race. Purdue University. https://www.education.purdue.edu/geri/new-publications/gifted-education-in-the-united-states

Gentry, M., Desmet, O. A., Karami, S., Lee, H., Green, C., Cress, A., Chowkase, A., & Gray, A. (2021). Gifted education's legacy of high stakes ability testing: Using measures for identification that perpetuate inequity. *Roeper Review, 43*(4), 242–255. https://doi.org/10.1080/02783193.2021.1967545

Goffney, I., Gutiérrez, R., & Boston, M. (2018). *Rehumanizing mathematics for black, indigenous, and Latinx students*. Reston: National Council of Teachers of Mathematics.

Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences, 99*(6), 4115–4120. https://doi.org/10.1073/pnas.06238189

Griggs, M. S., Rimm-Kaufman, S. E., Merritt, E. G., & Patton, C. L. (2013). The responsive classroom approach and fifth grade students' math and science anxiety and self-efficacy. *School Psychology Quarterly, 28*(4), 360–373. https://doi.org/10.1037/spq0000026

Hanushek, E. A., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2019). The achievement gap fails to close: Half century of testing shows persistent divide between haves and have-nots. *Education next, 19*(3), 8–17.

Harper, S. R. (2010). An anti-deficit achievement framework for research on students of color in STEM. *New Directions for Institutional Research, 2010*(148), 63–74. https://doi.org/10.1002/ir.362

Havard, B., Nguyen, G. N., & Otto, B. (2018). The impact of technology use and teacher professional development on US national assessment of educational progress (NAEP) mathematics achievement. *Education and Information Technologies, 23*(5), 1897–1918. https://doi.org/10.1007/s10639-018-9696-4

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Hickman, C. J., & Sherman, H. J. (2019). *Learning mathematics successfully: Raising self-efficacy in students, teachers, and parents*. Charlotte: Information Age Publishing.

Hiller, S. E., Kitsantas, A., Cheema, J. E., & Poulou, M. (2022). Mathematics anxiety and self-efficacy as predictors of mathematics literacy. *International Journal of Mathematical Education in Science and Technology, 53*(8), 2133–2151. https://doi.org/10.1080/0020739X.2020.1868589

Johnsen, S. K., & Sheffield, L. J. (2021). *Using the common core state standards for mathematics with gifted and advanced learners*. Routledge.

Jungert, T., & Andersson, U. (2013). Self-efficacy beliefs in mathematics, native language literacy and foreign language amongst boys and girls with and without mathematic difficulties. *Scandinavian Journal of Educational Research, 57*(1), 1–15. https://doi.org/10.1080/00313831.2011.621140

Kalaycioğlu, D. B. (2015). The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Educational Sciences Theory and Practice, 15*(5), 1391–1401.

Keşan, C., & Kaya, D. (2018). Mathematics and science self-efficacy resources as the predictor of academic success. *International Online Journal of Educational Sciences, 10*(2), 45–58. https://doi.org/10.15345/iojes.2018.02.004

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics, 22*(2), 310–339. https://doi.org/10.1177/1932202X1102200206

Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43. https://doi.org/10.1007/s10648-010-9141-8

Kotok, S. (2017). Unfulfilled potential: High-achieving minority students and the high school achievement gap in math. *High School Journal, 100*(3), 183–202. https://doi.org/10.1353/hsj.2017.0007

Larsen, N. E., & Jang, E. E. (2022). Instructional practices, students' self-efficacy and math achievement: A multi-level factor score path analysis. *Canadian Journal of Science, Mathematics and Technology Education, 21*(4), 803–823. https://doi.org/10.1007/s42330-021-00181-3

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209–231. https://doi.org/10.3102/0162373711431604

Lubienski, S. T. (2006). Examining instruction, achievement, and equity with NAEP mathematics data. *Education Policy Analysis Archives, 14*(14), 1–33. https://doi.org/10.14507/epaa.v14n14.2006

Lubienski, S. T., & Lubienski, C. (2006). School sector and academic achievement: A multilevel analysis of NAEP mathematics data. *American Educational Research Journal, 43*(4), 651–698. https://doi.org/10.3102/00028312043004651

McCoach, D. B. (2010). Hierarchical linear modeling. In G. R. Hancock, G. R. Hancock, R. O. Mueller, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 123–140). Routledge.

McConney, A., & Perry, L. B. (2010). Science and mathematics achievement in Australia: The role of school socioeconomic composition in educational equity and effectiveness. *International Journal of Science and Mathematics Education, 8*(3), 429–452. https://doi.org/10.1007/s10763-010-9197-4

Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: Impact of intraclass correlation and sample size on type-I error. *Frontiers in Psychology, 2*, 1–6. https://doi.org/10.3389/fpsyg.2011.00074

National Center for Education Statistics. (2009). NAEP technical documentation item scaling models. https://nces.ed.gov/nationsreportcard/tdw/analysis/scaling_models.aspx

National Center for Education Statistics. (2016). NAEP technical documentation: Plausible values versus individual scores. https://nces.ed.gov/nationsreportcard/tdw/analysis/est_pv_individual.asp

National Center for Education Statistics. (2022b). NAEP technical documentation: NAEP assessment IRT parameters. https://nces.ed.gov/nationsreportcard/tdw/analysis/scaling_irt.aspx

National Center for Education Statistics. (2022a). NAEP technical documentation NAEP assessment sample design. https://nces.ed.gov/nationsreportcard/tdw/sample_design/

National Assessment Governing Board. (2020). *Mathematics framework for the 2019 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.

Pitsia, V., Biggart, A., & Karakolidis, A. (2017). The role of students' self-beliefs, motivation and attitudes in predicting mathematics achievement: A multilevel analysis of the Programme for International Student Assessment data. *Learning and Individual Differences, 55*, 163–173. https://doi.org/10.1016/j.lindif.2017.03.014

Plucker, J. A., Hardesty, J., & Burroughs, N. (2013). *Talent on the sidelines: Excellence gaps and America's persistent talent underclass*. Stamford: University of Connecticut.

Plucker, J. A., Peters, S. J., & Schmalensee, S. (2017). Reducing excellence gaps: A research-based model. *Gifted Child Today, 40*(4), 245–250. https://doi.org/10.1177/1076217517723949

Polat, N., Zarecky-Hodge, A., & Schreiber, J. B. (2016). Academic growth trajectories of ELLs in NAEP data: The case of fourth-and eighth-grade ELLs and non-ELLs on mathematics and reading tests. *The Journal of Educational Research, 109*(5), 541–553. https://doi.org/10.1080/00220671.2014.993461

Raudenbush, S.W., & Congdon, R.T. (2021). *HLM 8: Hierarchical linear and nonlinear modeling*. Scientific Software International.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Newcastle upon Tyne: Sage.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Rodriguez, D., Carrasquillo, A., Garcia, E., & Howitt, D. (2022). Factors that challenge English learners and increase their dropout rates: Recommendations from the field. *International Journal of Bilingual Education and Bilingualism, 25*(3), 878–894. https://doi.org/10.1080/13670050.2020.1722059

Sakellariou, C. (2022). The reciprocal relationship between mathematics self-efficacy and mathematics performance in US high school students: Instrumental variables estimates and gender differences. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.941253

Sandilos, L. E., Baroody, A. E., Rimm-Kaufman, S. E., & Merritt, E. G. (2020). English learners' achievement in mathematics and science: Examining the role of self-efficacy. *Journal of School Psychology, 79*, 1–15. https://doi.org/10.1016/j.jsp.2020.02.002

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347–367. https://doi.org/10.1177/1094428107308906

Schunk, D. H., & DiBenedetto, M. K. (2021). Self-efficacy and human motivation. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 153–179). Amsterdam: Elsevier. https://doi.org/10.1016/bs.adms.2020.10.001

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 15–31). Academic Press.

Schwery, D. A. (2015). *How do mathematics self-efficacy and gender interact to predict mathematics achievement in fifth through eighth graders?* (Publication No. 3723935) [Doctoral dissertation, University of South Dakota]. ProQuest Dissertations and Theses database.

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology, 87*(3), 386–398. https://doi.org/10.1037/0022-0663.87.3.386

Soland, J., & Sandilos, L. E. (2021). English language learners, self-efficacy, and the achievement gap: Understanding the relationship between academic and social-emotional growth. *Journal of Education for Students Placed at Risk, 26*(1), 20–44. https://doi.org/10.1080/10824669.2020.1787171

Solomon, D., Battistich, V., & Hom, A. (1996). Teacher beliefs and practices in schools serving communities that differ in socioeconomic level. *The Journal of Experimental Education, 64*(4), 327–347.

Stapleton, L. M. (2013). Using multilevel structural equation modeling techniques with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 521–562). Charlotte: Information Age Publishing.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Stiff, L. V., Johnson, J. L., Akos, P., Tate, W. F., King, K. D., & Anderson, C. R. (2011). Examining what we know for sure: Tracking in middle grades mathematics. In W. F. Tate, K. D. King, & C. R. Anderson (Eds.), *Disrupting tradition: Research and practice pathways in mathematics education* (pp. 63–77). National Council of Teachers of Mathematics.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2021). The talent development megamodel: A domain-specific conceptual framework based on the psychology of high performance. In R. J. Sternberg & D. Ambrose (Eds.),

Conceptions of giftedness and talent (pp. 425–442). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-56869-6_24

Terman, L. M. (1925). *Genetic studies of genius Vol. 1. Mental and physical traits of a thousand gifted children*. Redwood City: Stanford University Press.

U.S. Department of Education & Institute of Education Sciences. (2005). IES Style Guide. https://nces.ed.gov/statprog/styleguide/pdf/styleguide.pdf

U.S. Department of Education (2018). *Improving basic programs operated by local educational agencies (Title I, part A)*. https://www2.ed.gov/programs/titleiparta/index.html#:~:text=Schools%20in%20which%20children%20from,of%20the%20lowest%2Dachieving%20students

Usher, E. L. (2009). Sources of middle school students' self-efficacy in mathematics: A qualitative investigation. *American Educational Research Journal, 46*(1), 275–314. https://doi.org/10.3102/0002831208324517

Van der Biji, J. J., & Shortridge-Baggett, L. M. (2001). The theory and measurement of the self-efficacy construct. *Scholarly Inquiry for Nursing Practice, 15*(3), 189–207.

Webster, B. J., & Fisher, D. L. (2010). Accounting for variation in science and mathematics achievement: A multilevel analysis of Australian data Trends in International Mathematics and Science Study (TIMSS). *School Effectiveness and School Improvement, 11*(3), 339–360. https://doi.org/10.1076/0924-3453(200009)11:3;1-G;FT339

Wiederkehr, V., Darnon, C., Chazal, S., Guimond, S., & Martinot, D. (2015). From social class to self-efficacy: Internalization of low social status pupils' school performance. *Social Psychology of Education, 18*(4), 769–784. https://doi.org/10.1007/s11218-015-9308-8

Williams, J. H. (2005). Cross-national variations in rural mathematics achievement. *Journal of Research in Rural Education, 20*(5), 1–18.

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology, 5*, 100075. https://doi.org/10.1016/j.metip.2021.100075

Wu, J. (2015). *A comprehensive analysis of the NAEP data from Native American youth concerning excellence gaps* (Publication No. 3734112) [Doctoral dissertation, Purdue University]. ProQuest Dissertations and Theses database.

Yang, Y., & Gentry, M. L. (2023). Striving to excel in STEM: Insights from underrepresented, minoritized graduate students with high academic ability. *Gifted Child Quarterly, 67*(2), 110–136. https://doi.org/10.1177/0016986222111

Yang, Y., & Maeda, Y. (2023). *An investigation of the excellence gaps in mathematics education: Evidence from NAEP 2019 [Manuscript in preparation]*. Purdue Univeristy.

Young, J. L., Young, J. R., & Ford, D. Y. (2017). Standing in the gaps: Examining the effects of early gifted education on Black girl achievement in STEM. *Journal of Advanced Academics, 28*(4), 290–312. https://doi.org/10.1177/1932202X17730549

Zhang, D., & Wang, C. (2020). The relationship between mathematics interest and mathematics achievement: Mediating roles of self-efficacy and mathematics anxiety. *International Journal of Educational Research, 104*, 101648. https://doi.org/10.1016/j.ijer.2020.101648

Publisher's Note