Journal of School Psychology 97 (2023) 123–151

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Probing gaps in educational outcomes within the U.S.: A dual moderation multiple mediator latent growth model☆

R. Sergio Guglielmi *

Lake Forest College, United States of America

ARTICLE INFO

Editor: Craig A. Albers
Action Editor: Danny Hajovsky

Keywords:
Moderated mediation
Calibration bias
Math achievement
Postsecondary attainment
Latent growth model
Longitudinal analyses

ABSTRACT

Racial/ethnic disparities in math achievement are especially troubling because math proficiency predicts long-term educational outcomes, but the mechanisms underlying these disparities remain unclear. Previous research has demonstrated that across diverse samples, both within and outside the United States, the relation between students' academic aspirations and later postsecondary attainment is mediated by initial levels of math ability and by growth in that ability across time. The key issue examined in this investigation is the extent to which students' underestimation or overestimation of their math ability (i.e., calibration bias) moderates those mediated effects and whether this moderation varies as a function of race/ethnicity. Using data from two longitudinal national surveys (i.e., NELS:88 and HSLS:09), these hypotheses were tested in samples of East Asian American, Mexican American, and Non-Hispanic White American high school students. In both studies and in all groups, the model explained large portions of the variance in post-secondary attainment. In East Asian Americans and non-Hispanic White Americans, calibration bias moderated the effect mediated by 9th grade math achievement. The strength of this effect was greatest at high levels of underconfidence and steadily weakened as self-confidence grew, suggesting that some degree of underconfidence may be achievement-promoting. Indeed, in the East Asian American sample, this effect became negative at high levels of overconfidence (i.e., academic aspirations actually predicted the lowest postsecondary attainment levels). Educational implications of these findings are discussed and possible reasons for the failure to find moderation effects in the Mexican American sample are explored.

1. Introduction

Comparative cross-national assessments, such as the Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), provide regular reminders that the U.S. math achievement gap is longstanding, seemingly intractable, and growing wider. The most recent PISA results confirm these sobering conclusions. For both boys and girls, the U.S. math mean was significantly below the average of the 37 participating members of the Organization for Economic Cooperation and Development (2019) and ranked seventh from the bottom. Even more troubling are conspicuous and stubborn math achievement gaps among racial/ethnic groups within the U.S. itself. Math performance is consistently highest in East Asian American students,

☆ Author Note: The author is grateful to Nancy Brekke for her careful reading of a draft and very constructive feedback.

* Corresponding author at: Department of Psychology, Lillard Science Center, , Lake Forest College, 555 North Sheridan Road, Lake Forest, IL 60045, United States of America.

E-mail address: guglielm@lakeforest.edu.

https://doi.org/10.1016/j.jsp.2023.01.001
Received 23 November 2021; Received in revised form 6 October 2022; Accepted 10 January 2023
Available online 15 February 2023
0022-4405/

lowest in Hispanic and Black American students, and intermediate in their Non-Hispanic White peers (U.S. Department of Education, 2020; Guglielmi, 2008, 2012; Guglielmi & Brekke, 2018).

Such disparities are worrisome because math performance is associated with important life outcomes. In both majority and minority samples, for instance, early math proficiency repeatedly has been found to predict college admission and completion (e.g., Davis-Kean et al., 2022; Gambini et al., 2022; Lee et al., 2008; Nomi et al., 2021; Perna & Titus, 2005; Schneider, 2003; Sciarra & Whitson, 2007; Thompson et al., 2006; Trusty & Niles, 2003). Although the correlational nature of that research precludes causal inferences, there is a growing number of quasi-experimental intervention studies that allow the estimation of causal effects when randomization is not feasible. Several of these studies point to the effectiveness of improving K–12 math skills in raising college admission and/or degree attainment rates, especially among minority and socioeconomically disadvantaged students (for a useful review, see Poulsen, 2019; and for recent empirical examples, see Nomi et al., 2021, and Xu et al., 2020).

The close connection between math proficiency and postsecondary outcomes may help explain racial/ethnic group disparities in educational attainment. Among people 25 years and older in the U.S., 64.8% of Asian Americans, 47.9% of Non-Hispanic Whites, and only 25.4% of Hispanics had earned at least an Associate's degree as of 2021 (U.S. Census Bureau, 2021). Bridging these gaps in postsecondary attainment requires a clearer insight into the mechanisms that underlie racial/ethnic differences in math achievement.

1.1. Understanding math achievement gaps

To a large extent, math achievement disparities reflect widespread racial, ethnic, and socioeconomic inequities. It is not surprising that the lowest performers in PISA and TIMMS international assessments are countries that lack the financial resources to invest as heavily in education as their wealthier counterparts, just as the lowest-performing racial/ethnic groups in the U.S. attend underfunded schools at disproportionate rates (Baker, 2021). Most recently, by widening many resource gaps (e.g., the digital divide), the Covid pandemic has also disproportionately impacted the education of economically disadvantaged groups (Cruz, 2021). Some scholars (e.g., Kendi, 2019) have argued that using the term "achievement gaps" to describe inter-group differences in academic performance is problematic, as it appears to highlight the shortcomings of individual students who belong to particular racial/ethnic groups, while seemingly ignoring the pernicious effects of systemic structural conditions that contribute to group disparities in educational outcomes. To eradicate the sociopolitical and institutional determinants of current educational inequities, it would be helpful to shift attention to the many "opportunity gaps" that have originated from a long history of discrimination, marginalization, and oppression (e.g., Flores, 2007; Mayor & Suarez, 2019; Quinn, 2020).

There is evidence, however, that even after removing the impact of sociostructural barriers (e.g., low socioeconomic status, poor access to resources), considerable portions of the variance in achievement remain unexplained (Heath et al., 2008; Jackson, 2012; Rothon, 2007). Psychological and cultural forces can substantially attenuate or exacerbate the damaging impact of socioeconomic disadvantage (Harwell et al., 2017; Kim et al., 2019; Liu et al., 2020). Moreover, whereas educators often have limited ability to enact broad structural changes, they can implement classroom interventions that address the psychological and cultural contributors to racial/ethnic achievement disparities. This is the focus of the present research. First, however, a brief overview of the theoretical framework that guides this investigation would be helpful.

1.2. Previous research: context-invariant, modifiable predictors of math achievement

Although group differences in math achievement have attracted the attention of many previous studies, one line of research is particularly relevant here. Guglielmi and Brekke (2017) noted that patterns of cross-national disparities in math achievement parallel the racial/ethnic disparities within the U.S. For example, East Asian students consistently outperform their international peers when they are assessed in their own countries, and immigrant samples of East Asian students outperform their classmates in U.S. schools despite the striking contextual changes associated with migration (e.g., language, culture, school system). The same consistency is evident in students from Mexico who tend to show low math achievement levels when assessed both in their own country and in U.S. schools. Similarly, non-Hispanic White students' achievement levels fall between the other two groups in both cross-national and intranational comparisons (Guglielmi & Brekke, 2017; U.S. Department of Education, 2020). To identify variables that might explain this stable pattern of group disparities, Guglielmi and Brekke (2017) proposed a math achievement and postsecondary attainment model that includes several literature-based, context-invariant, and modifiable predictors: parental aspirations, student aspirations, belief in the value of effort, homework time, and academic self-perception.

This model was empirically tested by Guglielmi and Brekke (2018) on cross-sectional comparative data from PISA-2012 and longitudinal U.S. data from the National Education Longitudinal Study and was then cross-validated on the more recent Education Longitudinal Study. In the analyses of PISA data, model fit was compared across top performing (East Asian), below average (U.S.), and very poorly performing (Mexico) countries. The context-invariance and transportability of those math performance predictors was then evaluated by fitting the model to longitudinal data from East Asian American, Mexican American, and Non-Hispanic White American students within U.S. high schools. The achievement predictors specified in the model were indeed found to be context-independent, time invariant, and transportable, as they could reliably forecast achievement levels in students assessed in their own country as well as in samples of students from immigrant families in the U.S. The same model of math achievement and postsecondary attainment exhibited outstanding fit in all samples, between and within countries, and across time (i.e., in both the older and more recent national longitudinal studies).

These findings have important implications for educational policy and practice as they point to a set of psychological variables that account for individual differences in math performance and might be worthwhile targets of educational interventions. At the same time, the findings failed to explain the large group differences in math achievement and postsecondary attainment levels that were found in that investigation and that consistently emerge from cross-national and intra-national comparisons. Viewed from this perspective, model specification was incomplete and was missing some important variable that could explain those group disparities in educational outcomes.

1.3. The present focus: what might moderate the relation between predictors and educational outcomes?

A sizeable body of work (some of this literature and its findings are reviewed below) on self-regulated learning has drawn attention to the importance of self-perception calibration accuracy (the degree of consistency between self-perceived ability and actual performance) in explaining individual and group differences in achievement. On the basis of this work, it seems plausible to hypothesize that the way predictors relate to educational outcomes varies as a function of calibration accuracy, and group differences in the way predictors relate to those outcomes surfaces only when this metacognitive monitoring moderator is included in the model.

1.3.1. Calibration of self-perception

A strong reciprocal association between academic self-perception (e.g., confidence, self-efficacy, self-concept) and academic outcomes is supported by a voluminous literature (for reviews, see Honicke & Broadbent, 2016; Huang, 2011; Richardson et al., 2012; Schneider & Preckel, 2017; Stankov, 2013; Talsma et al., 2018; Unrau et al., 2018). However, a more nuanced analysis of this relation indicates that the direction and extent to which self-beliefs predict scholastic performance depend on how well calibrated those beliefs are. A frequently reported finding is that students' calibration accuracy is generally quite poor and is biased in the direction of ability overestimation (e.g., Bol & Hacker, 2001; Ehrlinger et al., 2008; Kuncel et al., 2005; Mihalca, 2014; Pajares & Graham, 1999; Schwartz & Beaver, 2015; Serra & DeMarree, 2016; Talsma et al., 2019). This overconfidence or self-enhancement bias (i.e., *positive miscalibration*) is typically associated with poor performance (e.g., Bol et al., 2005; Hacker et al., 2008; Kruger & Dunning, 1999; Miller & Geraci, 2011; Osterhage et al., 2019; Serra & DeMarree, 2016; Zimmerman et al., 2017). Often referred to as the Dunning-Kruger effect, the combination of poor skills and overconfidence casts a "double curse," whereby the affected individual's ineptitude leads to errors and to the inability to recognize those errors, which then inevitably hampers self-improvement efforts (Dunning, 2011; Dunning et al., 2003). Although the literature cited above generally finds that accurate calibration (i.e., good correspondence between perceived and actual ability) is predictive of good performance, there is also evidence that *negative miscalibration* (i.e., ability underestimation) is associated with favorable outcomes.

Chiu and Klassen (2010), for example, conducted secondary analyses of PISA data and found that both accurate calibration and negative miscalibration predicted higher math scores than positive miscalibration. Similarly, Talsma et al. (2019) sought to determine the optimal level of academic self-efficacy for good academic performance in a sample of 207 undergraduates. Consistent with the general findings reported in the literature, high self-efficacy levels were, in fact, associated with good performance. However, when calibration of self-beliefs relative to achievement outcomes was examined, negative miscalibration (i.e., ability underestimation) was found to be more predictive of good performance than either positive miscalibration or accurate calibration, neither of which was significantly associated with performance. These findings, and similar ones (e.g., Callender et al., 2016; Kruger & Dunning, 1999), have important implications for educational policy and practice: widespread recommendations for raising students' self-efficacy beliefs without considering the degree to which these beliefs align with reality might actually hinder academic performance. Some underconfidence and self-doubt may instead be motivating and achievement-promoting.

Such findings may also help explain the results of Guglielmi and Brekke's (2018) study. Consistent with the broader self-efficacy literature, in all groups, math self-concept was significantly related to math achievement, but the large racial/ethnic disparities in math achievement and postsecondary attainment remained unexplained. It is argued here that the general association between academic self-perception and achievement does not capture group differences in metacognitive regulatory mechanisms that may, in fact, explain group disparities in educational outcomes. The important message to be drawn from these findings is that correlations between self-efficacy and achievement, so often reported in the literature, would be more informative if the moderating role of self-belief calibration were taken into account. High self-efficacy participants, for example, may still underrate their own competence relative to their actual performance levels. More specifically, even when East Asian students' academic self-perception is more positive than that of their Western peers, their achievement levels may nevertheless exceed their self-perceived ability and this negative miscalibration may then contribute to a strong performance. East Asian students in U.S. classrooms may perceive themselves as "big fish" in a pond of lower math achievement and although this may boost their self-efficacy beliefs, these beliefs may still underestimate their actual ability. Measures of calibration bias are needed to assess the moderating contribution of self-perception accuracy in the prediction of educational outcomes.

1.3.2. Calibration measures

To assess the extent to which different levels of calibration accuracy influence educational endpoints, appropriate calibration measures need to be included in the model. Two frequently used measures (see Schraw, 2009, for a helpful review) are *absolute accuracy* and *bias index*. In both cases, confidence judgments taken before (prediction) or after (postdiction) task performance are compared to the actual outcome, often by subtracting one from the other. When absolute accuracy is assessed, however, the result of the subtraction is squared and the direction of judgment errors is therefore removed. This yields an overall measure of precision; that is, how well perception and reality agree, regardless of the direction of deviations from perfect agreement. Conversely, the bias index is computed without squaring the subtraction value; therefore, this signed difference retains information about both the magnitude of deviations from perfect agreement as well as the direction of those deviations (with positive values indicating overconfidence and negative values reflecting underconfidence).

Freeman et al. (2017) compared the performance of these two calibration measures in the prediction of various academic outcomes and found that the bias index was the most predictive and informative measure. Similarly, in the relatively recent study by Talsma et al. (2019) described earlier, hierarchical regression analyses indicated that an absolute accuracy measure of metacognitive monitoring failed to predict actual performance (i.e., grades on written assignments and exams), whereas a bias index was strongly predictive, with negatively miscalibrated (i.e., underconfident) participants outperforming their accurately calibrated and positively miscalibrated peers.

In the study by Guglielmi and Brekke (2018) described earlier, student aspirations were significant predictors of math achievement and postsecondary attainment, after the influence of gender, SES, and parental aspirations was controlled. The direct and indirect effects of student aspirations on postsecondary attainment were large and statistically significant in all racial/ethnic groups and in both datasets analyzed (NELS and ELS). Of the specific indirect effects, the largest, and again significant in all groups and in both datasets, involved a mediational chain that included student aspirations → math achievement intercept → postsecondary attainment. Left unexplained, however, were large group differences in educational outcomes that emerged from the analyses of both datasets. The present study addressed this issue by testing a dual moderation multiple mediator model in which the extent to which math achievement mediates the relation between student aspirations and postsecondary attainment depends on math self-perception calibration bias and, in turn, the extent to which math self-perception calibration bias moderates that relation depends on race/ethnicity, the second moderator.

1.3.3. *Purpose of this investigation and hypotheses tested*

Based on the literature reviewed earlier, it was predicted that the mediating role of math achievement would be stronger in accurately calibrated and negatively miscalibrated (i.e., underconfident) students than in their positively miscalibrated (i.e., overconfident) peers. Furthermore, the moderating influence of math self-perception calibration was expected to vary as a function of race/ethnicity. The modesty bias characteristic of East Asian cultures (Cai et al., 2011; Eaton & Dembo, 1997; Fu et al., 2016; Klassen, 2004) supports the prediction that East Asian American students would be more negatively miscalibrated than their Mexican American or Non-Hispanic White American counterparts. More specifically, it was hypothesized that, at equal levels of educational aspirations, students who overestimate their math ability would have lower levels of math performance than their more modest counterparts and this underachievement would then undermine their postsecondary educational prospects. Moreover, to the extent that calibration of self-perceived math competence varies across racial/ethnic groups, these group differences in self-perception could then help explain differences in educational outcomes.

The doubly moderated (i.e., moderated moderated) multiple mediator model that captures these hypotheses is shown in Fig. 1. Hayes' (2015, 2018a, 2018b) useful recommendations for conceptualizing and testing such models are adopted in the present research. As shown in Fig. 1, both math achievement growth factors (intercept and slope) were hypothesized to mediate the relation between student aspirations and postsecondary attainment. In each case, the moderated path connects student aspirations to the mediator (i.e.,

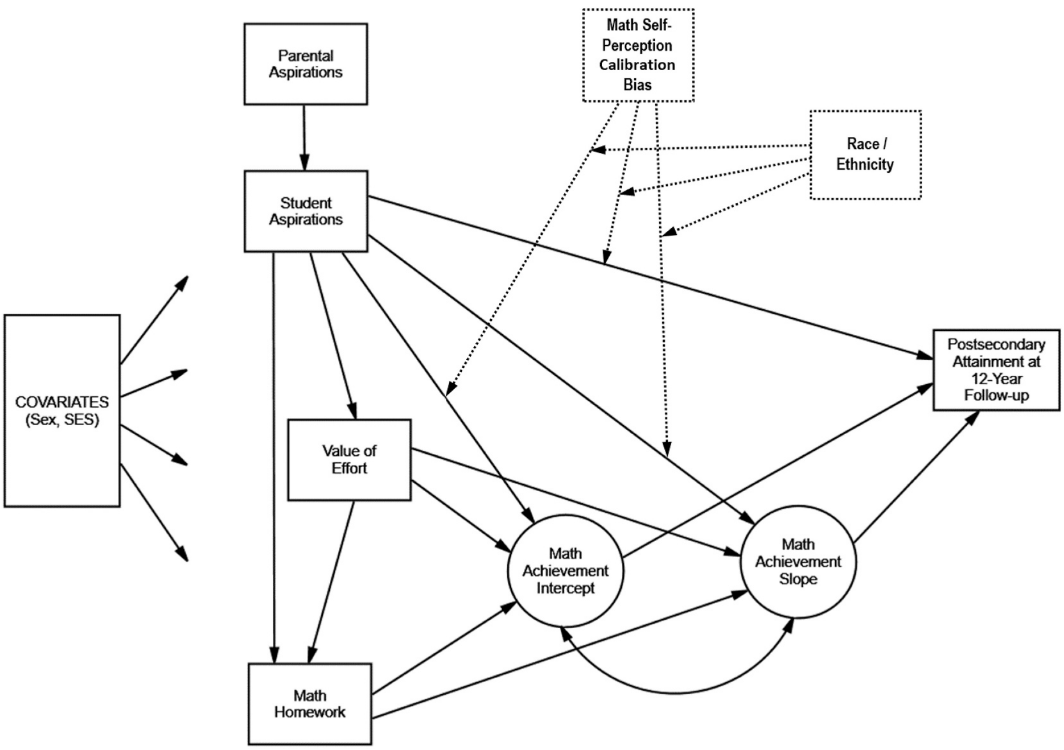

Fig. 1. Dual moderation multiple mediator latent growth model.

first stage moderation; see Edwards & Lambert, 2007). In other words, the strength of the relation between student aspirations and math achievement growth factors was expected to vary as a function of math self-perception calibration bias. The contribution of the second moderator (race/ethnicity) can then be evaluated by suitable group comparisons. Evidence of group differences in the extent to which the predictor-mediator relations are moderated by calibration bias can provide valuable insights into possible mechanisms underlying math achievement disparities and the racial/ethnic gaps that are consistently found in postsecondary attainment levels.

2. General method

The multiple mediator model diagrammed in Fig. 1 builds on the model tested previously by Guglielmi and Brekke (2018) by examining specifically the moderating roles of math self-perception calibration and race/ethnicity. This model was fit to the data of three U.S. groups with markedly different levels of math performance and postsecondary attainment: (a) East Asian Americans, (b) Mexican Americans, and (c) Non-Hispanic White Americans. In two separate studies, longitudinal data from large nationally representative student samples made it possible to evaluate the hypothesis that group differences in math achievement and postsecondary attainment levels vary as a function of differences in math self-perception calibration bias. In Study 1, these issues were examined with data from an older dataset, the National Education Longitudinal Study (NELS 1988–2000). The robustness and reproducibility of the NELS findings were then assessed by fitting the same model to data from the more recent High School Longitudinal Study (HSLS 2009–2016). In the analyses of both datasets, variables were recoded, when needed, so that high values would reflect high levels of the construct. Also, variables with large variances were rescaled by dividing them by a constant to prevent model convergence problems that can occur when constructs are measured on very different scales (see Muthén & Muthén, 1998–2017). However, mean values reported in the Results section are given in the original metric.

2.1. Analytic strategy

All models were estimated with M*plus* 8.4 (Muthén & Muthén, 1998–2017). Given the complex survey design of both NELS and HSLS, model estimation included cluster, stratum, and weight variables needed to obtain standard errors and goodness-of-fit indices that recognize the non-independence of observations associated with the stratified, multistage, clustered structure of the data. The MLR estimator used in the analyses provides full information maximum likelihood (FIML) parameter estimates with robust standard errors and chi-square statistics even with non-normal incomplete data. Model estimation also included auxiliary variables as missing data correlates, an approach that has been consistently found to increase statistical power, improve precision and reduce bias of parameter estimates, and enhance the plausibility of the missing at random assumption (e.g., Eekhout et al., 2015; Lang & Little, 2018; Raykov & West, 2016; Rioux & Little, 2021; Yuan & Savalei, 2014; Yuan & Zhang, 2012).

A much-debated issue is whether maximum likelihood (ML) estimation should be used with ordinal data. Several simulation studies have compared the performance of an ML estimator and a weighted least squares (WLS) estimator, as well as their "robustified" variants (ML with robust standard errors [MLR in M*plus*] and weighted least square with mean and variance adjusted [WLSMV]), under various distributional conditions, sample sizes, and number of response categories. Generally, these studies have found that each estimation method has advantages and disadvantages (Chen et al., 2020; Koziol & Bovaird, 2018; Lei, 2009; Lei & Shiverdecker, 2020; Li, 2016; Suh, 2015). The usefulness of ML estimation becomes especially evident under certain conditions. When categorization generates large deviations from normality, for example, estimators that are robust to violations of distributional assumptions (i.e., MLR) yield unbiased parameter estimates and standard errors (Chou et al., 1991; DiStefano, 2002; Green et al., 1997; Lei, 2009; Liu & Thompson, 2021; Suh, 2015). In simulation studies, MLR estimation of models with missing ordinal data has proved superior to a robust weighted least square alternative (e.g., Chen et al., 2020; Jia & Wu, 2019), and the behavior of the robust ML estimator improves as the number of response categories increases above two or three (Bentler & Chou, 1987; Coenders et al., 1997). In fact, with five or six response categories, robust ML estimation performs as well as or better than WLS alternatives (Beauducel & Herzberg, 2006; Finney & DiStefano, 2013; Lei, 2009; Padgett & Morgan, 2021; Rhemtulla et al., 2012). Across the two datasets analyzed in this investigation, two variables (one in each dataset) were assessed with four response categories, and all the remaining ordinal variables included at least six and as many as 10 categories (see the Method section for details). Lastly, from a practical perspective, Gottfredson et al. (2009) pointed out that, compared to the full information robust ML estimator available in M*plus* (MLR), the robust limited information counterpart (WLSMV) uses pairwise deletion in the presence of missing data, does not estimate models with auxiliary variables, and requires the unrealistic assumption that data are missing completely at random.

2.2. The moderated multiple mediator latent growth model

In the present investigation, the focal interest was on a portion of the full model shown in Fig. 1. The specific hypothesis tested in the analyses of both datasets was that student aspirations predict future educational attainment through the mediation of high school growth trajectories in math achievement. It was also hypothesized that the student aspirations → math achievement path is moderated by the degree of calibration bias in students' perception of their math competence. Finally, it was expected that the moderating influence of math self-perception calibration would depend on participants' race/ethnicity, the second moderator.

These hypotheses are incorporated into a parallel multiple mediator model with first stage moderation that is conceptually shown in Fig. 2; its statistical representation, which was applied to the analyses of both datasets, is given in Fig. 3. The indirect effects of student aspirations (X) on postsecondary attainment (Y) through the mediation of the two math achievement growth factors ($M1$ and $M2$) were expected to vary across levels of calibration bias (W). The model displayed in Fig. 3 can be represented by the following equations:

$$M1 = i_{M1} + a_{11}X + a_{21}W + a_{31}XW + e_{M1}$$

$$M2 = i_{M2} + a_{12}X + a_{22}W + a_{32}XW + e_{M2}$$

$$Y = i_Y + c'_1X + c'_2W + c'_3XW + b_1M1 + b_2M2 + e_Y$$

The paths $X \rightarrow M1$, $X \rightarrow M2$, and $X \rightarrow Y$, were hypothesized to be moderated by W. In each of the above equations, the regression coefficient for the interaction of X and W (a_{31}, a_{32}, *and* c'_3, respectively), if significantly different from zero, would indicate that the path is indeed moderated. The specific moderated indirect effects of X on Y through the two mediators are.

$$\text{Via } M1 : (a_{11} + a_{31}W)^*b_1 = a_{11}b_1 + a_{31}b_1W$$

$$\text{Via } M2 : (a_{12} + a_{32}W)^*b_2 = a_{12}b_2 + a_{32}b_2W$$

In both cases, the moderated paths connect the predictor to the mediators (i.e., first stage moderation). A useful measure of the extent to which mediated effects are moderated was introduced by Hayes (2015), who referred to it as the *index of moderated mediation*. The weight of the moderator W in the function that connects X to Y through $M1$ and $M2$ is the product of the two component paths ($a_{31}b_1$ for $M1$ and $a_{32}b_2$ for $M2$). This weight is the index of moderated mediation and it indicates the amount of change in the indirect effect of X on Y through the mediator as W changes by one unit. If the index is significantly different from zero, it can be concluded that the mediated effects of X on Y is moderated by W (i.e., is conditioned on values of W). This means that if the indirect effect of student aspiration on postsecondary attainment through either of the two achievement mediators is found to be moderated by self-perception calibration bias, it would not make much sense to ask whether this moderated indirect effect is significantly different from zero because that question cannot be answered. The value of the conditional indirect effect is not fixed but will vary as a function of the moderator. Thus, once evidence of moderation is established, a probe needs to be conducted to determine how the magnitude of the conditional indirect effect changes at different values of the moderator (Hayes, 2015, 2018b; Preacher et al., 2006, 2007).

Several procedures are available for probing and visualizing moderated effects (for useful reviews, see Bauer & Curran, 2005; Hayes & Matthes, 2009; McCabe et al., 2018). With the *pick-a-point* approach, some values of the moderator are selected (usually the mean and one standard deviation above and below the mean) and the statistical significance of the indirect effect at those values is assessed. A common criticism of the pick-a-point approach is that the selection of moderator values is an arbitrary process that provides limited information. Spiller et al. (2013) aptly described it as a spotlight analysis. By contrast, the *Johnson-Neyman* (J-N) technique provides a floodlight test of moderation (Spiller et al., 2013); it is a more informative approach because the significance of the conditional indirect effect is plotted against all values of the moderator and confidence intervals can be constructed to define the "regions of significance" (i.e., the points across the full distribution of the moderator at which the conditional mediated effect is different from zero). This approach is illustrated in the Results sections below.

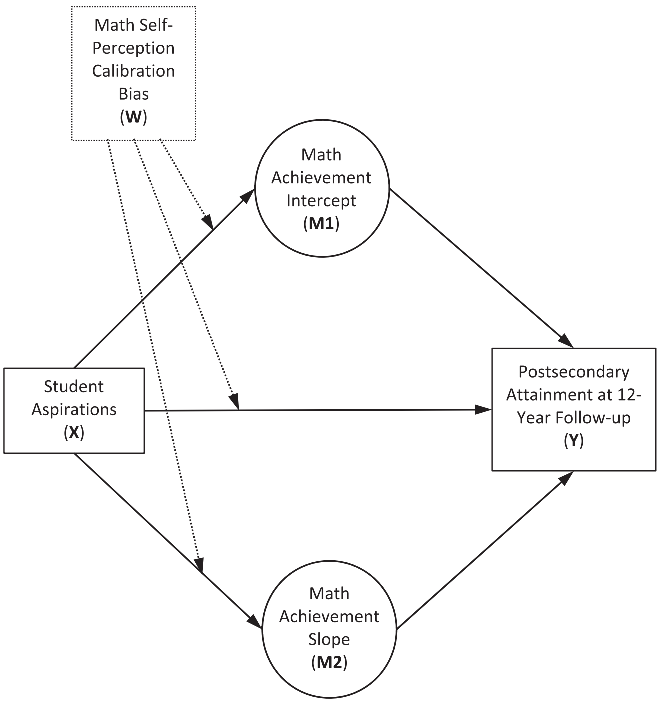

Fig. 2. Latent growth multiple mediator model with first stage moderation – conceptual model.

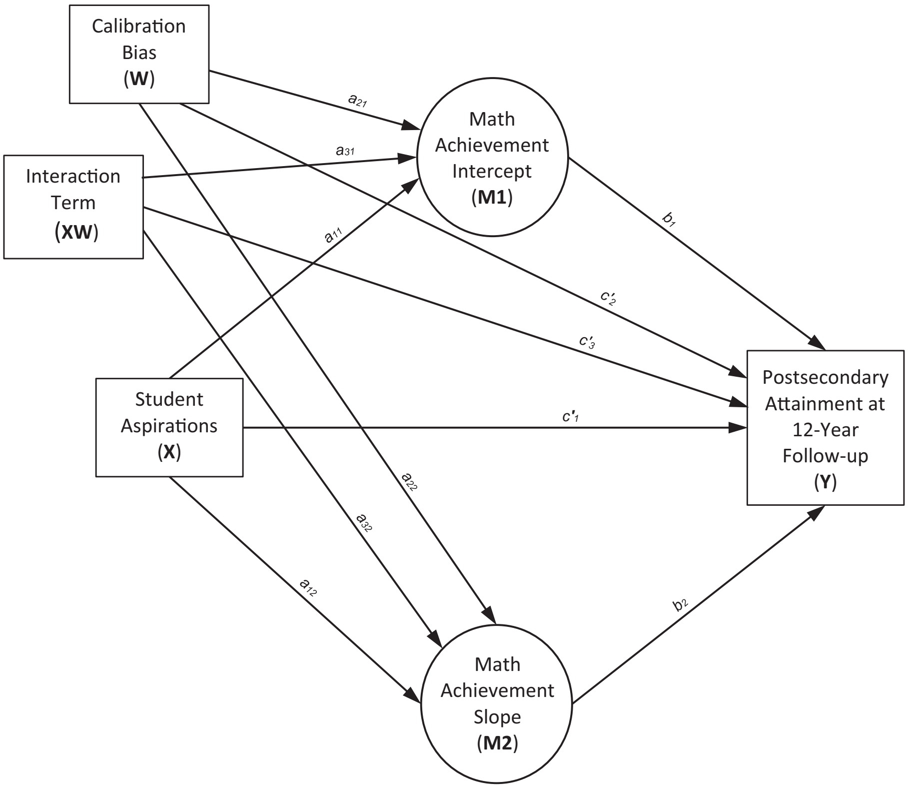

Fig. 3. Latent growth multiple mediator model with first stage moderation – statistical model.

Although in the current investigation the theoretical interest and analytic focus were on the conditional indirect effects represented in Figs. 2 and 3, all moderated mediation analyses were conducted within the full model, controlling for the influence of all the other variables shown in Fig. 1. There are conceptual and analytic advantages to simultaneously estimating all mediated and moderated effects in the context of the full structural model rather than in isolation (Bauer et al., 2006; Fairchild & MacKinnon, 2009). Furthermore, Hayes and Preacher (2013) recommended that, even when the contribution of direct effects might be of little theoretical relevance and the focal interest is moderation of mediated effects (as in the present investigation), direct effects be included in the estimated model. They argued that omission of these direct paths (i.e., constraining them to zero) could produce biased estimates of the moderated indirect effects. Thus, the moderated mediation model shown in Fig. 3 also includes a conditional direct effect from student aspirations to educational attainment.

In the present research, self-perception calibration bias is not construed as a transient state but rather as a stable individual disposition that is, to some degree, culturally shaped. This conceptualization of the construct is reflected in its operationalization. As explained more fully in the Method sections, in the analyses of both datasets, measures of self-perceived math ability were averaged across two waves of data; transcript-based math GPA, also averaged across two waves, was then subtracted from the self-perception measure. Positive values indicate ability overestimation, negative values indicate ability underestimation, and values close to zero reflect accurately calibrated math self-perception.

3. Study 1: NELS:88/2000

A nationally representative U.S. sample of almost 28,000 eighth-graders was followed periodically over the course of 12 years. The final follow-up occurred about 8 years after most sample members had graduated from high school. Full information about sample characteristics can be found in Hafner et al. (1990). The study's three waves of data (including assessments of math achievement), which were collected when the participants were in school (Grades 8, 10, and 12), make it possible to examine the extent to which math achievement growth trajectories mediated the relation between Grade 8 student academic aspirations and post-secondary attainment levels assessed 12 years later. Moreover, the extent to which bias in self-perceived math ability and race/ethnicity moderated those mediated effects could also be evaluated.

3.1. Method

3.1.1. Samples

NELS sample members were selected for analysis if they had participated in all waves of data collection, transcript data were available for them, and they belonged to one of the three racial/ethnic subgroups described earlier. These inclusion criteria yielded a total sample of approximately 8220 participants: Non-Hispanic White Americans ($N \approx 7150$), Mexican Americans ($N \approx 790$), and East Asian Americans ($N \approx 270$). The East Asian American sample included ≈ 150 Chinese students, ≈ 80 South Korean students, and ≈ 40 Japanese students. (Please note that samples sizes are approximate because the IES Data Security Office of the U.S. Department of Education requires restricted-use data licensees to round to the nearest 10 all unweighted sample size numbers. Also note that a license for restricted-use datasets can be requested from that office: https://nces.ed.gov/statprog/licenseapp/LicenseInfo).

3.1.2. Measures

The dataset includes all the variables needed to test the research hypotheses outlined above.

Sampling Weight. F4TRSCWT is the appropriate sampling weight in longitudinal analyses that include respondents who participated in all data collection waves and for whom transcript data are available.

Observed Covariates. F2SEX is the most complete indicator of participants' gender available in the dataset. Also provided in the dataset is F2SES1, a continuous measure of student socioeconomic status constructed from standardized estimates of mother's and father's educational level and occupation, and family income.

Parental Aspirations. Eighth graders rated on a 6-point scale (from *less than high school* to *a postgraduate degree*) how far in school each parent would like them to go. Mother's and father's aspirations (BYS48A and BYS48B) were averaged to construct a Time 1 measure of parental aspirations that has excellent reliability ($\omega = 0.913$).

Student Aspirations. Participants rated, on the same 6-point scale as Parental Aspirations, how far in school they thought they would get (BYS45).

Belief in the Value of Effort. Using a 4-point scale, participants rated, at Time 1, their level of agreement (from *strongly disagree* to *strongly agree*) with the statement "Good luck is more important than hard work" (BYS44C).

Math Homework. This Time 1 measure consists of students' estimates, on an 8-point scale (from *none* to *10 hours or more*), the number of hours per week they typically spent on math homework at Time 1 (BYS79A).

Math Self-Perception Calibration Bias. The dataset provides an 8th grade measure of self-reported math grades "from sixth grade up till now" (BYS81B), and a 10th grade measure of self-reported math grades "from the beginning of ninth grade until now" (F1S39A). In both cases, the range of self-reported math grades was from *Mostly below D* to *Mostly A*. Students' actual grades in all math courses taken in Grades 9 and 10 were obtained from the transcript files and the average math GPA was then subtracted from the averaged self-reported math grades to yield an index of calibration bias. Prior to subtraction, however, the variables were standardized to bring them to the same scale. They were then grand-mean centered so that, following subtraction, a value of zero would indicate perfect calibration accuracy. Again, negative values on this calibration index reflect ability underestimation (i.e., self-reported math grades lower than actual math GPA) and positive values indicate ability overestimation (i.e., self-reported math grades higher than actual math GPA).

Math Achievement. Standardized math tests were administered in Grades 8, 10, and 12 (Times 1, 2, and 3). Of the various types of test scores available in the dataset, those generated by Item Response Theory (IRT) are comparable across waves and are therefore well-suited for longitudinal analyses that focus on achievement gains across time. The psychometric properties of the NELS cognitive test scores are clearly documented in a report by Rock and Pollack (1995), which estimates reliabilities of math test scores as ranging from 0.89 to 0.94 for the three waves of the study.

Postsecondary Attainment. The final follow-up, 8 years after high school graduation, provides a 6-point measure of the highest level of postsecondary education attained: 0 = *no postsecondary education*; 1 = *some postsecondary education but no degree attained*; 2 = *attained an associate's degree*; 3 = *attained a bachelor's degree*; 4 = *attained a master's degree but not* higher; 5 = *attained a Ph.D. or professional degree*.

3.2. Results

The first analytic step was to determine, for the full sample and for each subgroup, the optimal shape of the change trajectory that most adequately described the repeated measures math achievement data (i.e., the latent growth curve). The remaining analyses zoomed in on the portion of the full model that was the specific focus of this investigation (Figs. 2 and 3). Two important questions were examined: (a) whether math achievement growth processes (intercept and slope) mediate the relation between 8th graders' educational aspirations and their post-secondary attainment levels, and (b) the extent to which those indirect effects are moderated by math self-perception calibration bias and by race/ethnicity.

3.2.1. Math achievement: latent growth model

Math achievement scores at three points in time were modeled as indicators of two latent growth factors: math achievement intercept and math achievement slope, which represent initial level of math achievement and developmental trajectories of change in math ability across the high school years. An unspecified nonlinear trajectory model in which the data themselves were used to identify the growth function that best represents change over time was found to be the best fitting model (see Chan, 1998; Curran & Hussong, 2003; Preacher et al., 2008). It should be noted that this is a saturated (just identified) model whose fit cannot be evaluated because these models will always fit to the data perfectly. Overidentification, however, was achieved after fixing the residual variance of the first math achievement indicator at a small value (0.001). When estimated on the full sample, this model exhibited outstanding fit, χ^2 (1, N ≈ 8210) = 2.421, $p = .120$, CFI = 1.000, TLI = 1.000, RMSEA = 0.013, 90% CI [0.000, 0.035], SRMR = 0.009. Multigroup invariance analyses indicated that the shape of this change trajectory was invariant across the three racial/ethnic groups. A chi-square difference test ($\Delta_{\chi 2}$) found that the fit of a model in which the Time 2 loading on the slope growth factor was freely estimated was not significantly degraded after constraining the same parameter to equality across groups, $\Delta_{\chi 2}$ (2) = 0.199, p = .905. In all groups, growth factor means and variances were significantly different from zero (all p-values < .001). Moreover, growth factor means were significantly different across groups (all p-values < .001), except for the mean math achievement slope in the Mexican vs. Non-Hispanic White comparison, Wald χ^2 (1) = 3.807, p = .051 (see Table 1).

3.2.2. Longitudinal moderated multiple mediator model

Despite a significant chi-square, known to be sensitive to sample size, the approximate fit indices gave evidence of very good fit when the full model was estimated in the whole sample, χ^2 (25, N ≈ 8210) = 61.625, p < .001, CFI = 0.998, TLI = 0.996, RMSEA = 0.013, 90% CI [0.009, 0.018], SRMR = 0.016. The model was then estimated separately in the three racial/ethnic groups. Both the chi-square goodness-of-fit and the approximate fit indices provided evidence of outstanding fit in the East Asian American sample, χ^2 (25, N ≈ 270) = 23.756, p = .533, CFI = 1.000, TLI = 1.000, RMSEA = 0.000, 90% CI [0.000, 0.045], SRMR = 0.038. The approximate fit indices indicated good fit in the Mexican American group, χ^2 (25, N ≈ 790) = 53.578, p < .001, CFI = 0.980, TLI = 0.948, RMSEA = 0.038, 90% CI [0.024, 0.052], SRMR = 0.044, and in the Non-Hispanic White American group, χ^2 (25, N ≈ 7150) = 61.622, p < .001, CFI = 0.998, TLI = 0.996, RMSEA = 0.014, 90% CI [0.010, 0.019], SRMR = 0.018. The set of variables in the model explained substantial portions of the variance in educational outcomes in all groups (see Table 2, which also provides R^2 information for Study 2). The remaining analyses reported here address the specific issue under examination in the present investigation, namely the moderated mediation processes that are hypothesized to shed light on group differences in educational outcomes.

Moderated Mediation Analyses. Table 1 shows, for each racial/ethnic group, the means of the variables included in the moderated mediation portion of the model as well as the significance of group differences in those means. Compared to the other two groups, East Asian American students had higher initial levels of math achievement and greater rates of improvement in math achievement across their high school years. East Asian American students also reached higher levels of postsecondary attainment, although the difference from the Non-Hispanic White sample is not statistically significant. Finally, East Asian American students were more negatively miscalibrated than the other two groups. Compared to East Asian Americans, Mexican American students were more positively miscalibrated and Non-Hispanic White Americans were more accurately calibrated. However, the difference in mean calibration bias between Mexican American and Non-Hispanic White American participants was not statistically significant. There were no significant group differences in student aspirations.

Parameter estimates and accompanying confidence intervals for the latent growth multiple mediator model with first stage moderation diagrammed in Fig. 3 are presented in Table 3. Again, this model was estimated not in isolation but within the full model shown in Fig. 1. Mediation analyses attempted to identify the specific mechanisms through which the student aspirations variable transmits its beneficial effects to postsecondary attainment. In the whole sample and in each racial/ethnic subsample, student aspirations significantly predicted initial levels of math achievement (a_{11} path in Fig. 3) which, in turn, significantly predicted postsecondary attainment (b_1). Similarly, except for the East Asian American sample, student aspirations significantly predicted the achievement slope mediator (a_{12}), which then significantly predicted postsecondary attainment (b_2).

The direct effect of student aspirations on postsecondary attainment (c'_1) was significant in the full sample and in the East Asian and Non-Hispanic White American subsamples, but not in the Mexican subsample. It should be noted that the inclusion of the c'_3 path makes this direct effect conditioned on calibration bias. If the c'_3 coefficient is found to be significantly different from zero, one can conclude that the direct effect is linearly dependent on (i.e., is moderated by) calibration bias. Table 3 shows evidence that this effect is moderated only in the Non-Hispanic White American sample. The findings of these analyses, however, hold little theoretical interest because the direct path was specified not on conceptual grounds, but to reduce bias in the estimation of the conditional indirect effects (Hayes, 2018b; Hayes & Preacher, 2013). The focal attention here is on these indirect effects and the key questions are whether the mediating role of math achievement growth factors varies for different levels of math self-perception calibration bias, and whether this moderation is dependent on race/ethnicity.

The parallel multiple mediator model with first-stage moderation shown in Fig. 3 specifies moderation of the paths from student aspirations to the two mediators (the two math achievement growth factors). Hayes' index of moderated mediation was estimated

Table 1
Study 1 – NELS. Moderated mediation model: means (and standard deviations) for three racial/ethnic groups.

Group	Student Aspirations	Math Achievement Intercept	Math Achievement Slope	Postsecondary Attainment	Calibration Bias
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
East Asian American	5.388_a (0.81)	44.624_a (12.502)	3.800_a (1.631)	2.364_a (1.537)	-0.168_a (0.65)
Mexican American	4.222_a (1.416)	29.159_b (8.59)	2.663_b (1.862)	1.047_b (0.942)	$.125_b$ (0.858)
NH White American	4.579_a (1.269)	37.298_c (11.863)	2.983_b (1.892)	1.595_a (1.236)	$.021_b$ (0.749)

Note. Wald test results for pairwise comparisons between groups. Within each column, means that share no subscript letters are significantly different. NH = Non-Hispanic.

Table 2
Percentage of variance in educational outcomes explained by the full set of predictors in Study 1 and Study 2 for the total sample and for each racial/ethnic group.

	Study 1: NELS				Study 2: HSLS			
Educational Outcome	Full Sample (N ≈ 8210)	East Asian American (*N* ≈ 270)	Mexican American (*N* ≈ 790)	NH White American (*N* ≈ 7150)	Full Sample (N ≈ 10,260)	East Asian American (*N* ≈ 450)	Mexican American (*N* ≈ 1100)	NH White American (*N* ≈ 8710)
Math Achievement Intercept	30.8	28.5	25.8	29.1	23.9	22.8	14.1	24.1
Math Achievement Slope	7.1	10.2	13.9	6.7	9.2	7.1	2.4	10.6
Postsecondary Attainment	42.6	40.7	22.3	43.1	45.0	45.5	26.3	45.5

Note. The IES Data Security Office of the U.S. Department of Education requires restricted-use data licensees to round to the nearest 10 all unweighted sample sizes. NH = Non-Hispanic.

separately for the two moderated mediation paths ($a_{31}b_1$ for *M1* and $a_{32}b_2$ for *M2*) in order to quantify the amount of change in the conditional indirect effect through each mediator for a one unit change in the moderator. Table 4 shows that in the full sample the index was significantly different from zero for the path mediated by *M1* and for the path mediated by *M2*. Thus, in the full sample, math self-perception calibration bias moderates the indirect effect of student aspirations on postsecondary attainment through initial levels of math achievement (*M1*), and through trajectories of change in math achievement (*M2*). The negative sign indicates that the indirect effect of student aspirations on postsecondary attainment via the two mediators is stronger in students who tend to be negatively miscalibrated (i.e., underestimate their math proficiency).

Conceptually interesting as this finding may be, it does not say anything about whether the moderating influence of calibration bias differs across racial/ethnic groups and whether such group differences might help explain disparities in educational achievement and attainment, the key questions under examination. To evaluate the extent to which race/ethnicity moderates the moderating influence of calibration bias on the indirect effect of student aspirations on postsecondary attainment through the mediation of initial math achievement, the index of moderated mediation was estimated separately in each group. Table 4 shows that this index was significantly different from zero, and again with a negative sign, in the East Asian American and in the Non-Hispanic White American subsamples, but not in the Mexican American group. Moreover, evidence of moderated mediation was significantly stronger in East Asian Americans than in both the Mexican American sample, Wald χ^2 (1) = 8.118, $p = .004$, and the Non-Hispanic White American sample, Wald χ^2 (1) = 7.281, $p = .007$, but the difference between the Mexican and Non-Hispanic White American samples was not significant, Wald χ^2 (1) = 2.391, $p = .122$. Conversely, the index of moderated mediation for the math achievement slope mediator was significant in the large full sample but was not significant in any of the three subgroups. Therefore, no further analyses were conducted in relation to this process.

Probing the Interaction. Evidence of significant moderation of the indirect effect via initial levels of math achievement calls for a probe aimed at determining how the magnitude and significance of the conditional indirect effect varies at different values of the moderator. Instead of choosing some arbitrary moderator values with the pick-a-point approach, the strength of the indirect effect was estimated for the full distribution of moderator values using the J-N technique described earlier. Incidentally, it should be noted that the results of the pick-a-point approach are embedded within the J-N plot and can be recovered from this plot by locating the magnitude and significance of the indirect effect for the specific values of the moderator chosen (e.g., mean ± 1 *SD*). Fig. 4 displays the results for the East Asian American sample. The conditional indirect effect of student aspirations on postsecondary attainment via the mediation of the math achievement intercept was plotted against all possible values of math self-perception calibration bias. The solid line shows the estimated conditional indirect effect, which is significant when the upper and lower 95% confidence intervals (dashed lines) do not include zero. As indicated by the negative slope, the conditional indirect effect progressively weakens as calibration bias moves away from negative values (i.e., underestimation of math ability) and is no longer significant when calibration bias reaches the zero point (i.e., accurate calibration). The indirect effect then becomes significant again when calibration bias values enter the positive region and are greater than 1.2, but now the effect is negative, which means that in this range of positive moderator values (i.e., overestimation of math ability) student aspirations are actually associated with lower levels of postsecondary attainment.

Fig. 5 displays the J-N plot for the Non-Hispanic White American sample. Please note, however, that the y-axis scale is not the same as for the East Asian American sample (Fig. 4) because using that scale would make it impossible for readers to discern at which moderator values the indirect effect falls within the region of significance. (The slopes of the three groups, plotted on the same scale, are displayed in Fig. 6). Fig. 5 shows that the conditional indirect effect remains significant up to moderator values lower than 2.08; this effect was no longer significant beyond this point and does not become negative. In other words, student aspirations predict Non-Hispanic White American students' postsecondary attainment at negative miscalibration (i.e., underconfidence) values but not at high levels of positive miscalibration (i.e., overconfidence).

A separate J-N plot for the Mexican American sample is not shown because the moderated mediation index was not significant in this group, and it does not make much sense to probe an interaction that is not there. Fig. 6, however, shows the slopes for the three groups on the same scale for ease of comparison, with the confidence bands removed to reduce clutter. These slopes represent the index of moderated mediation for each group. Although slopes for all three groups are negative, only the slopes of the East Asian American and the Non-Hispanic White American samples are significantly different from zero. The nearly flat slope of the Mexican American sample in Fig. 6 provides visual confirmation that the indirect effect is not moderated by calibration bias. The Non-Hispanic White American slope is also relatively flat but the statistical power accorded by the very large sample size helps make the moderated

Table 3

Study 1 – NELS. Latent growth multiple mediator model with first stage moderation. Unstandardized (and standardized) estimates and 95% confidence intervals by race/ethnicity.

	Full Sample		Asian American		Mexican American		NH White American	
Path	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Student Aspirations → Postsecondary Attainment (c'_1)	0.150*** (0.158)	0.122, 0.178	0.219** (0.155)	0.054, 0.383	0.012 (0.018)	−0.054, 0.078	0.162*** (0.167)	0.134, 0.191
Calibration Bias → Postsecondary Attainment (c'_2)	0.006 (0.004)	−0.123, 0.136	−0.529 (−0.301)	−2.102, 1.045	−0.280* (−0.255)	−0.548, −0.012	0.046 (0.028)	−0.083, 0.176
Student Aspirations x Calibration Bias → Postsecondary Attainment (c'_3)	−0.027 (−0.076)	−0.058, 0.004	0.076 (0.235)	−0.217, 0.359	0.047 (0.178)	−0.014, 0.108	−0.036* (−0.098)	−0.068, −0.004
Calibration Bias → Math Achievement Intercept (a_{21})	0.230** (0.147)	0.058, 0.402	3.104** (1.619)	1.032, 5.177	0.122 (0.123)	−0.133, 0.377	0.219* (0.139)	0.028, 0.410
Calibration Bias → Math Achievement Slope (a_{22})	0.012 (0.048)	−0.019, 0.043	0.086 (0.351)	−0.221, 0.393	0.034 (0.158)	−0.045, 0.112	0.011 (0.043)	−0.023, 0.045
Student Aspirations → Math Achievement Intercept (a_{11})	0.247*** (0.268)	0.217, 0.276	0.192* (0.125)	0.002, 0.382	0.181*** (0.299)	0.122, 0.240	0.256*** (0.276)	0.224, 0.288
Student Aspirations → Math Achievement Slope (a_{12})	0.029*** (0.194)	0.022, 0.035	0.011 (0.054)	−0.023, 0.045	0.037*** (0.281)	0.016, 0.057	0.027*** (0.182)	0.020, 0.034
Student Aspirations x Calibration Bias → Math Achievement Intercept (a_{31})	−0.076*** (−0.219)	−0.117, −0.035	−0.616** (−1.743)	−1.009, −0.222	−0.021 (−0.88)	−0.087, 0.045	−0.075** (−0.215)	−0.120, −0.030
Student Aspirations x Calibration Bias → Math Achievement Slope (a_{32})	−0.005 (−0.090)	−0.012, 0.002	−0.01 (−0.213)	−0.066, 0.047	−0.003 (−0.065)	−0.027, 0.020	−0.005 (−0.094)	−0.013, 0.002
Math Achievement Intercept → Postsecondary Attainment (b_1)	0.284*** (0.275)	0.247, 0.321	0.493*** (0.539)	0.363, 0.624	0.236*** (0.214)	0.135, 0.337	0.284*** (0.272)	0.246, 0.322
Math Achievement Slope → Postsecondary Attainment (b_2)	1.274*** (0.197)	1.106, 1.442	1.763** (0.246)	0.737, 2.789	1.180*** (0.231)	0.722, 1.639	1.283*** (0.196)	1.105, 1.461

Note. *$p < .05$, **$p < .01$, ***$p < .001$. NH = Non-Hispanic.

Table 4

Study 1 – NELS. Index of moderated mediation for two conditional indirect effects of student aspirations on postsecondary attainment by race/ethnicity. Unstandardized estimates and 95% confidence intervals.

	Full Sample		Asian American		Mexican American		NH White American	
Path	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Student Aspirations x Calibration Bias → Math Achievement Intercept → Postsecondary Attainment ($a_{31}b_{1}$)	−0.022***	−0.034, −0.010	-0.304_{a}**	−0.509, −0.099	$-.005_{b}$	−0.021, 0.011	-0.021_{b}**	−0.035, −0.008
Student Aspirations x Calibration Bias → Math Achievement Slope → Postsecondary Attainment ($a_{32}b_{2}$)	−0.006***	−0.015, 0.002	-0.017_{a}	−0.118, 0.084	$-.004_{a}$	−0.031, 0.023	-0.007_{a}	−0.016, 0.003

Note. *$p < .05$, **$p < .01$, ***$p < .001$. Wald test results for pairwise comparisons between groups. Within each row, estimates that share no subscript letters are significantly different. NH = Non-Hispanic.

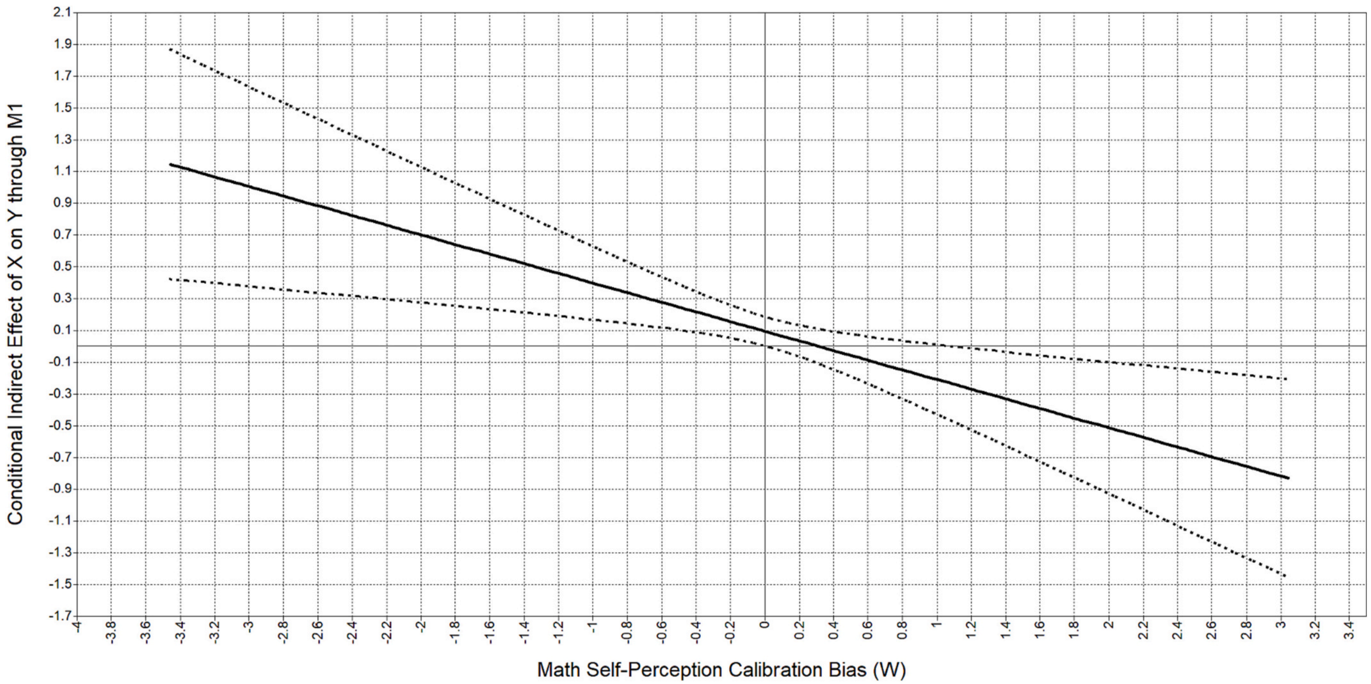

Fig. 4. Study 1 – NELS. Johnson-Neyman probe of moderation of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias in the East Asian American sample.
Note. Solid line = conditional indirect effect, dotted lines = upper and lower 95% confidence intervals.

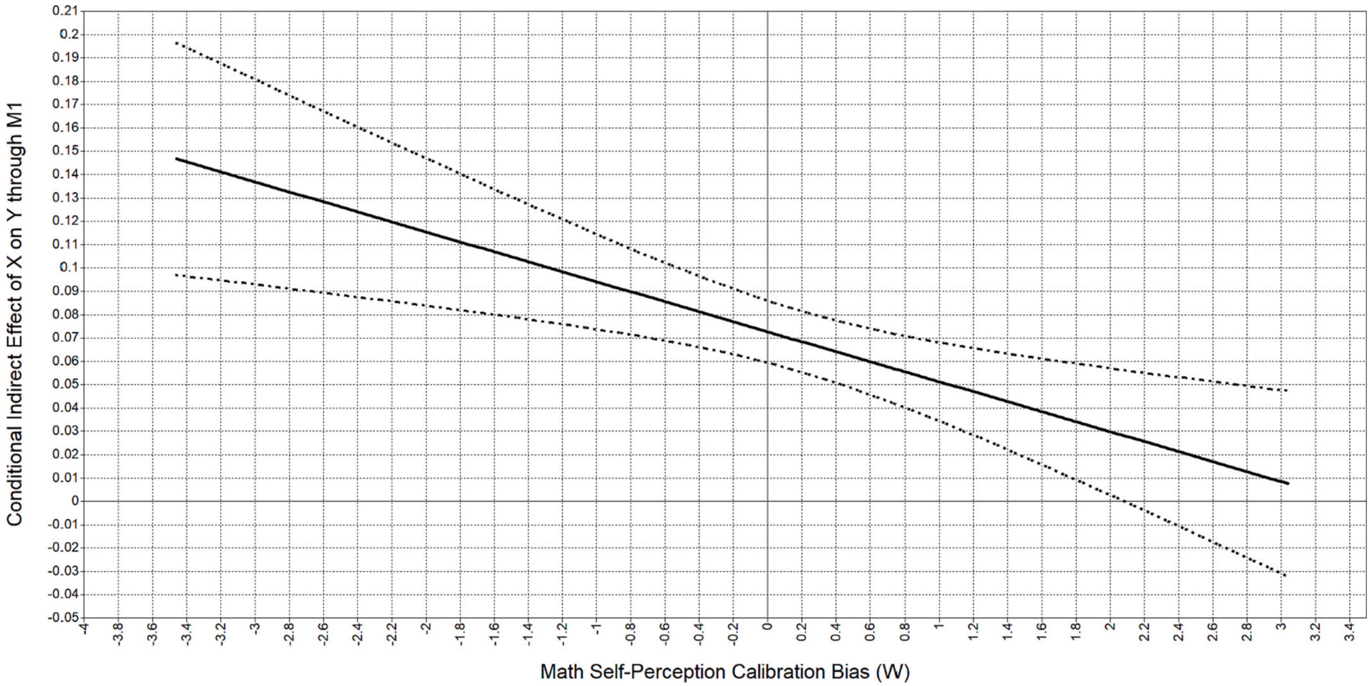

Fig. 5. Study 1 – NELS. Johnson-Neyman probe of moderation of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias in the NH-White American sample.
Note. Solid line = conditional indirect effect, dotted lines = upper and lower 95% confidence intervals.

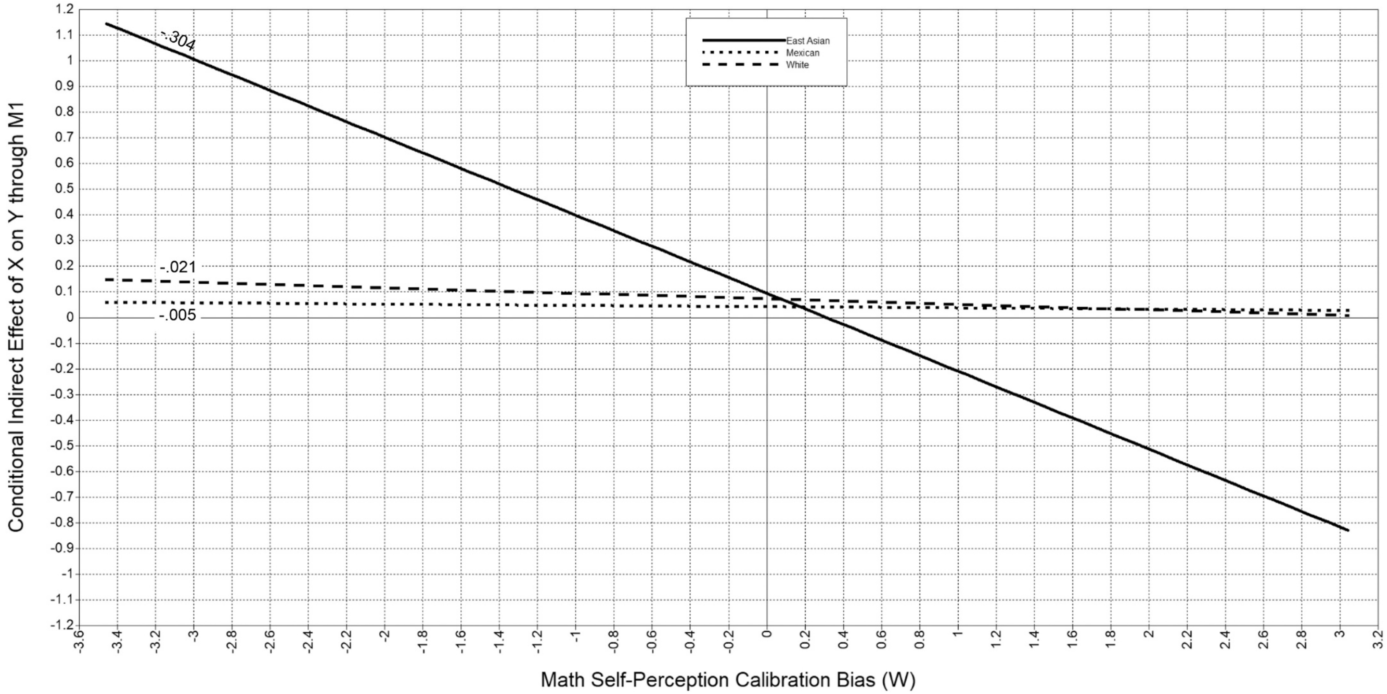

Fig. 6. Study 1 – NELS. Slopes of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias: comparison of three racial/ethnic groups.

mediation index statistically significant. By contrast, the slope for the East Asian American sample shows large deviations from the zero line and indicates that the strength of the conditional indirect effect is negatively related to math self-perception calibration bias, such that the greater students' underconfidence, the more strongly their aspirations predict levels of postsecondary attainment through the mediation of initial math achievement. As calibration then moves from the accurate range to growing levels of overconfidence, student aspirations predict increasingly lower levels of postsecondary attainment via the *M1* mediator. More specifically, as shown in Table 4 and in Fig. 6, the index of moderated mediation is −0.304 for the East Asian American sample, which means that as calibration of math self-perception increases by one unit, the indirect effect of student aspirations on postsecondary attainment through the mediation of math achievement intercept decreases by 0.304 unit.

3.3. Summary of Study 1 findings

In all three groups, an unspecified nonlinear growth function best captured the intraindividual change in math achievement across 4 school years. When the latent growth model shown in Fig. 2 was embedded in the full longitudinal model shown in Fig. 1, good fit was found in all three groups and the set of predictors accounted for substantial amounts of variance in postsecondary attainment (i.e., the distal outcome). Except for students' educational aspirations, group differences in mean values were evident for each of the variables included in the moderated mediation analyses. Assessment of the first stage moderated mediation model (shown in Fig. 3) indicated that, in the full sample and in the East Asian and Non-Hispanic White subsamples, the strength of the indirect effect of student aspirations on postsecondary attainment mediated by *M1* was a function of math self-perception calibration bias. Conversely, similar analyses for the second mediator (growth in math achievement) showed that the moderated mediation index was not significant in any group.

4. Study 2: HSLS:09

HSLS:09 is a more recent longitudinal study sponsored by the National Center for Education Statistics. In the 2009–2010 school year, a nationally representative sample of more than 23,000 9th graders participated in the base-year data collection. Follow-ups took place in 2012, when most sample members were finishing the 11th grade, and in 2016, when they had joined the workforce or were pursuing postsecondary education. A summary of sample characteristics can be found in LoGerfo et al. (2011).

4.1. Method

4.1.1. Samples

Only students who participated in all three waves of the study were included in the analyses. Availability of transcript data and membership in one of the three racial/ethnic subgroups under examination in this project were additional selection criteria. The resulting sample of approximately 10,260 participants included ≈ 8710 Non-Hispanic White Americans, ≈ 1100 Mexican Americans, and ≈ 450 East Asian Americans (≈ 280 Chinese, ≈ 170 Japanese or Korean). A license for this restricted-use dataset can be requested from https://nces.ed.gov/statprog/licenseapp/LicenseInfo.

4.1.2. Measures

All variables needed to test the research hypotheses are available in the dataset, which also provides statistically imputed estimates of missing values.

Sampling Weight. Of the multiple sample weights available in the dataset, W3W1W2STUTR is the most appropriate for longitudinal analyses when sample members participated in all three waves of the study and for whom transcript data were available.

Observed Covariates. X1SEX identifies participant gender, and X4X2SES is a composite measure of socioeconomic status based on the responding parent/guardian's education, occupation, and family income.

Parental Aspirations. X1PAREDEXPCT is a base-year measure of the highest level of education parents expected their children would reach, using a 10-point scale (from *less than high school* to *completion of PhD/MD/Law/or other professional degree*).

Student Aspirations. Ninth graders used the same 10-point scale as was used in Parental Aspirations to indicate how far in school they expected to get (X1STUEDEXPCT).

Effort. A base-year variable conceptually analogous to the NELS measure is not available in this dataset. As a best approximation, an effort variable was created as the mean of S1LATE, S1NOHWDN, S1NOPAPER, and S1NOBOOKS, which assess how often the participant goes to class late, or without homework done, or without a pencil and paper, or without books. This measure has adequate reliability (coefficient $\alpha = 0.683$).

Math Homework. Ninth graders estimated how many hours on a typical school day they spent on math homework, using a 6-point scale, from *less than one hour* to *five or more hours* (S1HRMHOMEWK).

Math Self-Perception Calibration Bias. A math self-efficacy measure was collected at base-year (X1MTHEFF) and again at the first follow-up (X2MTHEFF). Both are standardized scales that assess participants' confidence in their ability to (a) do an excellent job on math tests, (b) understand the math textbook, (c) master skills taught in the math course, and (d) do an excellent job on math assignments. Both scales have outstanding reliability, Cronbach's alpha = 0.90 and 0.89, respectively (Ingels et al., 2011, 2013). Following the procedure described in Study 1, the standardized math self-efficacy measures at the two points in time were averaged. Actual math GPAs in Grade 9 and in Grade 11 were calculated from transcript-based grades in all math courses taken in each of those two grades. The two math GPAs were averaged, standardized, and grand-mean centered. A calibration bias index was then constructed by subtracting this average from the standardized, centered, and averaged math self-efficacy measure.

Math Achievement. Standardized math assessments were conducted when students were in Grade 9 and again in Grade 11. The variable X1TXMSCR is the IRT-generated base-year estimate of number of correct responses, and X2TXMSCR is the analogous variable collected at the first follow-up. For both math assessments, the IRT-estimated reliability is 0.92 (Ingels et al., 2011, 2013).

Postsecondary Attainment. The second follow-up data, collected 3 years after high school graduation, provide a measure of institutional selectivity of the first postsecondary institution attended, if any (X4PS1SELECT). This variable was recoded to include a "not enrolled" category and to assess on a 7-point scale both the type of institution attended (2-years vs. 4-years) and its selectivity based on the 2010 Carnegie Classification (Carnegie Foundation, 2011). Possible outcomes range from *not enrolled* to *highly selective 4-year institution*.

4.2. Results

In Study 2, the model shown in Fig. 1 was tested with data from the HSLS:09 to cross-validate the results of Study 1 and determine the extent to which the construct relations that emerged from the first study were invariant across samples and across time. The HSLS provides only two waves of math achievement data. Although an examination of change that occurs between two points in time is often somewhat imprecisely characterized as "growth modeling," the functional form of the math "growth" trajectory could not be evaluated, as change between two points in time can be represented only by a straight-line function. Moreover, univariate two-wave panel designs yield "growth" models that are statistically underidentified. Parameters of theoretical interest, however, could be estimated after fixing the residual variances of the two math achievement indicators at 0.001 and, when embedded in the full moderated multiple mediator model, this measure of change made it possible to examine whether initial levels of math achievement and change in those levels between Grade 9 and Grade 11 mediated the effect of student aspirations on postsecondary attainment and whether calibration bias moderated that effect differently in the three racial/ethnic groups. These are the key issues addressed below.

4.2.1. Longitudinal moderated multiple mediator model

Fit of the full model was acceptable, although not as good as in Study 1, especially for the Mexican American subsample. In the whole sample, χ^2 (15, N ≈ 10,260) = 362.153, $p < .001$, CFI = 0.975, TLI = 0.907, RMSEA = 0.047, 90% CI [0.043, 0.052], SRMR = 0.049. In the East Asian American subsample, χ^2 (15, N ≈ 450) = 22.378, $p = .098$, CFI = 0.982, TLI = 0.933, RMSEA = 0.033, 90% CI [0.000, 0.060], SRMR = 0.083. In the Mexican American group, χ^2 (15, N ≈ 1100) = 78.550, $p < .001$, CFI = 0.927, TLI = 0.773, RMSEA = 0.062, 90% CI [0.049, 0.076], SRMR = 0.050. In the Non-Hispanic White American group, χ^2 (15, N ≈ 8710) = 409.105, $p < .001$, CFI = 0.977, TLI = 0.917, RMSEA = 0.055, 90% CI [0.050, 0.060], SRMR = 0.051. As in Study 1, the set of predictors included in the model explained a substantial amount of variance in postsecondary attainment (see Table 2).

Moderated Mediation Analyses. The means of the variables that capture the moderated mediation hypotheses under examination and the significance of group differences in those means are shown in Table 5. As in Study 1, East Asian American participants had significantly higher means than the other two groups on the two math achievement growth factors (intercept and slope). They also had higher levels of postsecondary attainment, except that their mean was again not significantly different from that of Non-Hispanic White Americans. As in Study 1, there were no significant group differences in educational aspirations. Finally, as in Study 1, on average, East Asian Americans were negatively miscalibrated, Mexican American were positively miscalibrated, and Non-Hispanic White Americans were more accurately calibrated. All group differences in calibration were statistically significant ($p < .001$).

Path coefficients for the moderated multiple mediator processes, estimated after controlling for the other variables in the full model, are reported for each racial/ethnic group in Table 6. With respect to the prediction of the postsecondary attainment outcome, the pattern of results across groups is remarkably similar to the one yielded by the analyses of NELS data. Paths from math achievement growth factors to postsecondary attainment were significant in all groups, and the direct path from student aspirations to postsecondary attainment was also significant in all groups except the Mexican American sample.

The same pattern of moderating effects was found across the two studies. The index of moderated mediation was computed for the conditional indirect effect of student aspirations on postsecondary attainment via initial levels of math achievement ($a_{31}b_1$) and via gains in achievement over time ($a_{32}b_2$). The results can be found in Table 7. As in Study 1, in the full sample this index was significantly different from zero for the path mediated by the math achievement intercept (*M1*), but not for the path mediated by the math

Table 5
Study 2 – HSLS. Moderated mediation model: means (and standard deviations) for three racial/ethnic groups.

Group	Student Aspirations	Math Achievement Intercept	Math Achievement Slope	Postsecondary Attainment	Calibration Bias
East Asian American	7.754_a (2.318)	52.028_a (10.573)	17.051_a (5.672)	4.466_a (2.039)	-0.538_a (0.847)
Mexican American	5.732_a (2.935)	37.501_b (10.447)	12.037_b (7.0)	1.922_b (1.783)	$.434_b$ (1.013)
NH White American	6.757_a (2.66)	41.745_c (11.475)	13.578_c (6.433)	3.080_a (2.219)	-0.082_c (0.975)

Note. Wald test results for pairwise comparisons between groups. Within each column, means that share no subscript letters are significantly different. NH = Non-Hispanic.

Table 6

Study 2 – HSLS. Latent growth multiple mediator model with first stage moderation. Unstandardized (and standardized) estimates and 95% confidence intervals by race/ethnicity.

	Full Sample		Asian American		Mexican American		NH White American	
Path	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Student Aspirations → Postsecondary Attainment (c'_1)	0.127*** (0.159)	0.096, 0.159	0.208*** (0.257)	0.104, 0.313	0.02 (0.033)	−0.083, 0.123	0.156*** (0.189)	0.134, 0.179
Calibration Bias → Postsecondary Attainment (c'_2)	−0.206* (−0.094)	−0.377, −0.035	−0.376 (−0.159)	−1.273, 0.521	−0.214 (−0.121)	−0.727, 0.300	−0.295*** (−0.130)	−0.428, −0.162
Student Aspirations x Calibration Bias → Postsecondary Attainment (c'_3)	−0.161* (−0.086)	−0.284, −0.039	0.095 (0.051)	−0.562, 0.752	−0.108 (−0.064)	−0.478, 0.262	−0.096 (−0.050)	−0.213, 0.020
Calibration Bias → Math Achievement Intercept (a_{21})	−0.061 (−0.053)	−0.142, 0.020	0.309 (0.25)	−0.063, 0.681	−0.19 (−0.186)	−0.408, 0.027	−0.054 (−0.046)	−0.139, 0.031
Calibration Bias → Math Achievement Slope (a_{22})	−0.05 (−0.076)	−0.105, 0.005	0.049 (0.074)	−0.194, 0.292	−0.114 (−0.166)	−0.266, 0.037	−0.044 (−0.068)	−0.098, 0.009
Student Aspirations → Math Achievement Intercept (a_{11})	0.108*** (0.258)	0.091, 0.124	0.041 (0.097)	−0.029, 0.110	0.046 (0.128)	−0.004, 0.095	0.122*** (0.285)	0.109, 0.135
Student Aspirations → Math Achievement Slope (a_{12})	0.036*** (0.152)	0.027, 0.046	0.053** (0.23)	0.019, 0.086	0.026 (0.107)	−0.006, 0.057	0.038*** (0.158)	0.031, 0.045
Student Aspirations x Calibration Bias → Math Achievement Intercept (a_{31})	−0.087* (−0.088)	−0.158, −0.017	−0.398** (−0.405)	−0.691, −0.105	−0.002 (−0.002)	−0.213, 0.210	−0.078* (−0.079)	−0.153, −0.004
Student Aspirations x Calibration Bias → Math Achievement Slope (a_{32})	−0.008 (−0.013)	−0.054, 0.039	−0.037 (−0.069)	−0.223, 0.150	0.085 (0.13)	−0.058, 0.228	−0.016 (−0.028)	−0.062, 0.030
Math Achievement Intercept → Postsecondary Attainment (b_1)	0.574*** (0.3)	0.504, 0.644	0.675*** (0.353)	0.488, 0.861	0.501*** (0.29)	0.264, 0.738	0.563*** (0.291)	0.516, 0.611
Math Achievement Slope → Postsecondary Attainment (b_2)	0.653*** (0.195)	0.551, 0.755	0.585* (0.165)	0.066, 1.105	0.630*** (0.245)	0.287, 0.973	0.647*** (0.187)	0.565, 0.729

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. NH = Non-Hispanic.

Table 7
Study 2 – HSLS. Index of moderated mediation for two conditional indirect effects of student aspirations on postsecondary attainment by race/ethnicity. Unstandardized estimates and 95% confidence intervals.

	Full Sample		Asian American		Mexican American		NH White American	
Path	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Student Aspirations x Calibration Bias → Math Achievement Intercept → Postsecondary Attainment ($a_{31}b_1$)	−0.050*	−0.090, −0.010	-0.269_a*	−0.474, −0.063	-0.001_b	−0.107, 0.105	-0.044_b*	−0.086, −0.002
Student Aspirations x Calibration Bias → Math Achievement Slope → Postsecondary Attainment ($a_{32}b_2$)	−0.005	−0.035, 0.025	-0.021_a	−0.137, 0.094	0.054_a	−0.049, 0.156	-0.010_a	−0.040, 0.020

Note. **p* < .05, ***p* < .01, ****p* < .001. Wald test results for pairwise comparisons between groups. Within each row, estimates that share no subscript letters are significantly different. NH = Non-Hispanic.

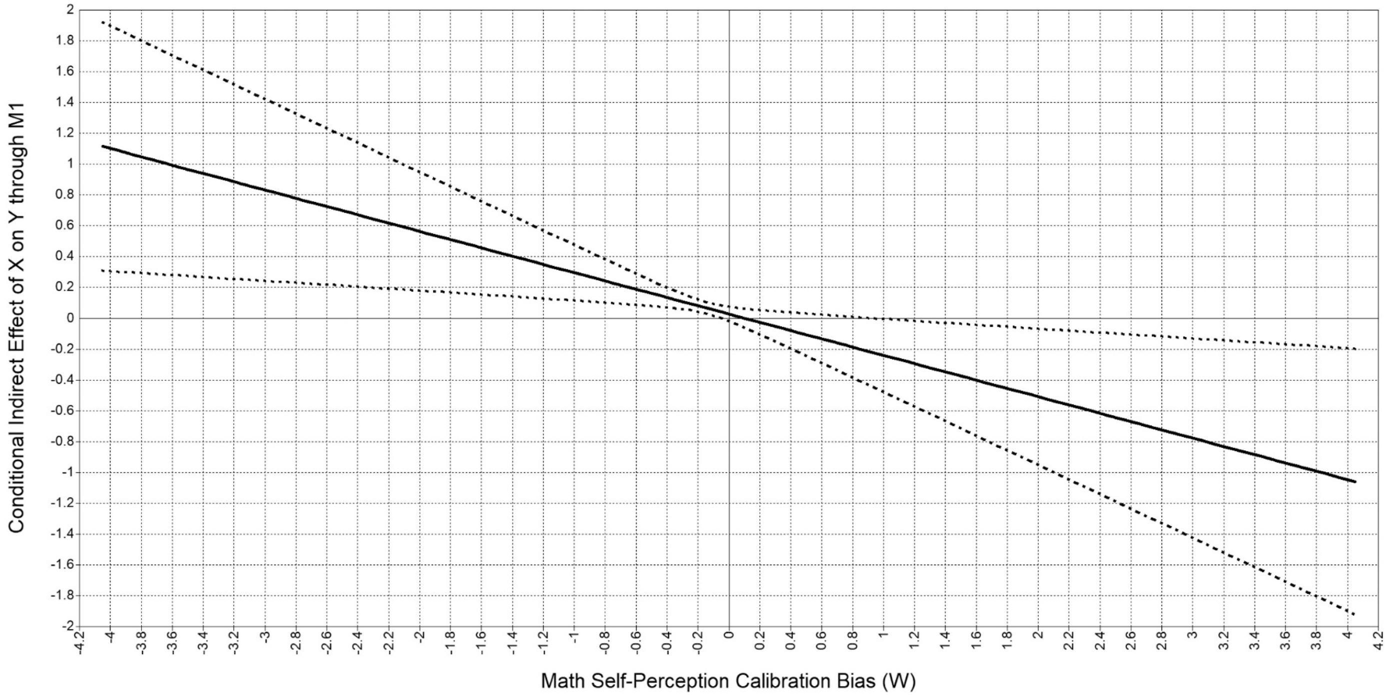

Fig. 7. Study 2 – HSLS. Johnson-Neyman probe of moderation of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias in the East Asian American sample.
Note. Solid line = conditional indirect effect, dotted lines = upper and lower 95% confidence intervals.

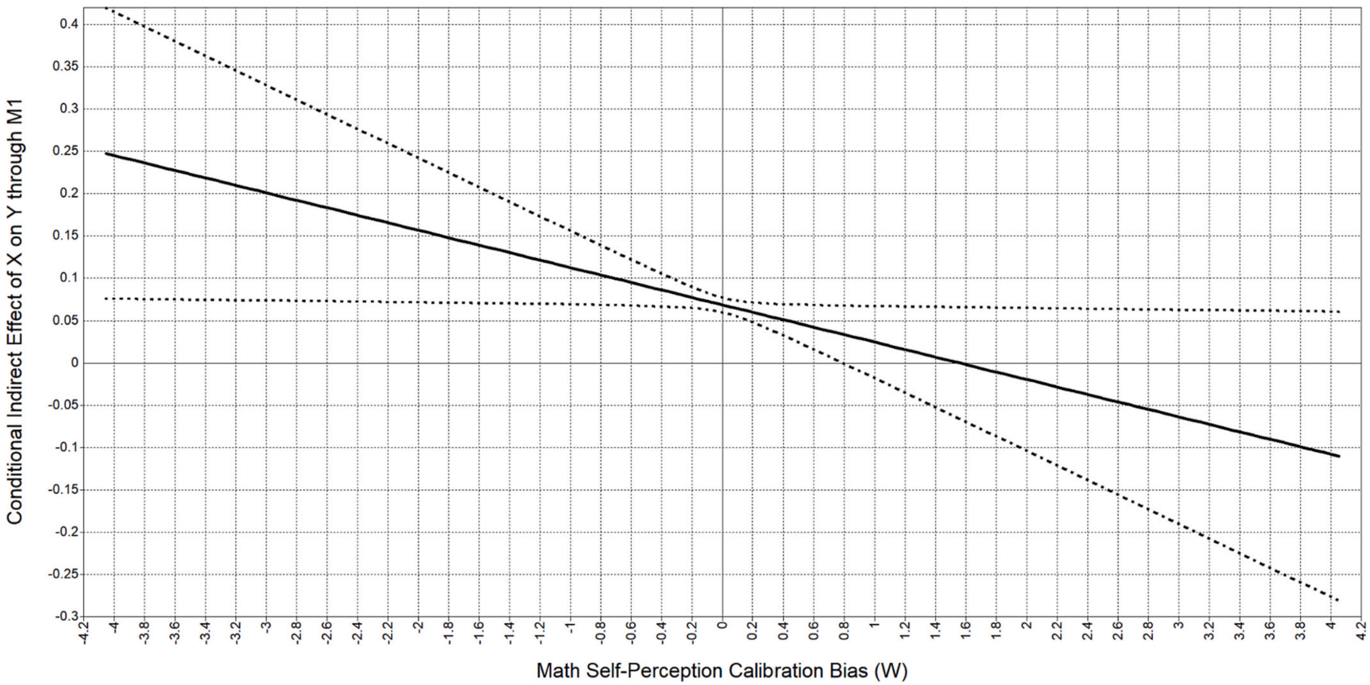

Fig. 8. Study 2 – HSLS. Johnson-Neyman probe of moderation of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias in the NH-White American Sample.
Note. Solid line = conditional indirect effect, dotted lines = upper and lower 95% confidence intervals.

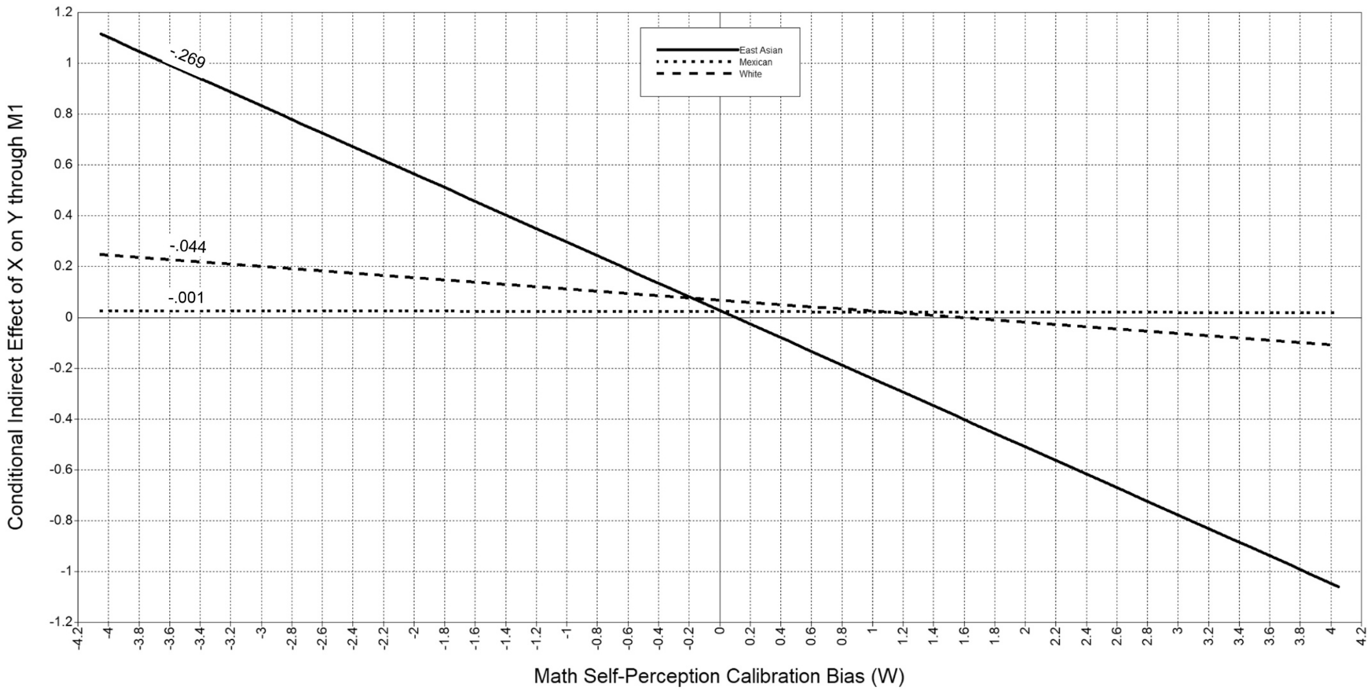

Fig. 9. Study 2 – HSLS. Slopes of the conditional indirect effect of student aspirations on postsecondary attainment through initial math achievement level by math self-perception calibration bias: comparison of three racial/ethnic groups.

achievement slope (*M2*). To determine whether the moderating effects of calibration bias differed across the three racial/ethnic groups, the index of moderated mediation was then estimated separately in each of those groups. As in Study 1, Table 7 shows that the index of moderated mediation through math achievement slope was not significant in any group or across groups. Conversely, for the path mediated by *M1*, as in Study 1 this index was significantly different from zero, and again with a negative sign, in the East Asian American and in the Non-Hispanic White American subsamples, but not significant in the Mexican American subsample. Furthermore, the index of moderated mediation had a larger negative value in the East Asian American group than in either the Mexican American sample, Wald χ^2 (1) = 5.236, $p = .022$, or the Non-Hispanic White American sample, Wald χ^2 (1) = 4.318, $p = .038$. As in Study 1, no significant difference was found between the latter two groups, Wald χ^2 (1) = 0.578, $p = .447$. It should be noted here that, in order to rule out the possibility that group differences in the strength of the indirect effect may have been present even without the moderation of calibration bias, an ancillary analysis was conducted in which the two specific indirect effects (one for each math achievement mediator) were estimated without the moderation of calibration bias. In both NELS and HSLS analyses, those unmoderated indirect effects were not significantly different between the East Asian American group and the other two samples. This confirms that, when group difference in educational outcomes are considered, a great deal of explanatory power would be lost if calibration bias were not included in the model.

Probing the Interaction. The Johnson-Neyman approach was used to determine whether the indirect effect of student aspirations on postsecondary attainment through the achievement intercept mediator (*M1*) varied in strength across the full range of calibration bias. Moreover, the hypothesis that these conditional indirect effects are dependent on race/ethnicity was tested by conducting the probe separately on each of the three groups. Fig. 7 provides the results for the East Asian American sample. As in Study 1, the steep negative slope shows that the conditional indirect effect is strongly positive and statistically significant at negative miscalibration values (underconfidence), is weakest and not significant when math self-perception is accurately calibrated, and grows increasingly negative and again statistically significant in the positive miscalibration region of the moderator distribution (overconfidence). Fig. 8 shows the J-N plot for the Non-Hispanic White American sample (note again that, for the reasons explained earlier, the y-axis scale is different from the one used in the East Asian American J-N plot). Consistent with the results of Study 1, the conditional indirect effect is positive and significant at all negative moderator values and up to positive values below 0.8, above which it is no longer statistically significant. Once more, the slope of the indirect effect is much flatter in this group than in the East Asian American sample, but group differences become fully apparent only after plotting the slopes for the three samples on the same scale (see Fig. 9). The similarity between the NELS and HSLS plots is striking. In both of them, the sharply negative slope in the East Asian American sample shows that the strength of the conditional indirect effect weakens as the moderator values move from negative to positive miscalibration. In fact, at high positive miscalibration levels, the sign of this conditional indirect effect again becomes negative, which indicates that, when math ability is overestimated, academic aspirations actually predict lower levels of postsecondary attainment.

4.3. Summary of Study 2 findings

Overall, the analyses of HSLS data replicated the pattern of NELS findings quite well. Table 2 shows that, in both studies, the full model explained more than 40% of the variance in postsecondary attainment in the East Asian American and Non-Hispanic White American samples, and more than 20% in the Mexican American sample. In both studies, there was no evidence of moderation for the indirect effect transmitted by math achievement slope, and in both studies group differences in mean levels of calibration bias followed the same pattern, as did the results of moderated mediation analyses and a probe of the moderated mediation effect.

5. General discussion

Past work by Guglielmi and Brekke (2018) indicated that, in all three racial/ethnic groups under consideration in the present research, parents' educational expectations for their children, students' own aspirations, belief in the value of effort, time spent on math homework, and math self-concept were important predictors of math achievement and, through it, of postsecondary attainment. Despite their interesting educational implications, however, these findings did not elucidate the mechanisms that may contribute to enduring group disparities in math achievement and postsecondary attainment. The main purpose of the present study was to investigate this issue by examining the possibility that the way predictors relate to educational outcomes depends on how accurately students perceive their own math ability. The self-regulated learning literature reviewed earlier suggests that the degree and direction of bias in self-perceived ability might help explain how the same predictors could forecast positive educational outcomes in all groups yet vary across groups in the strength of their beneficial effects. The results of Study 1 were cross-validated in Study 2, and the hypotheses guiding this investigation were generally supported. At least in some racial/ethnic groups, calibration of self-perceived math competence influenced the important link between educational aspirations and initial levels of math proficiency, a reliable and strong predictor of college admission and completion. More specifically, in East Asian American and, to a lesser extent, Non-Hispanic White American high school students, math ability underestimation was significantly associated with greater math achievement and, through it, with higher levels of postsecondary attainment. The strength of this negative indirect effect was found to be largest in the East Asian American sample and significantly different from the other two groups.

5.1. Implications of findings

Although inferences from non-experimental data require much caution, when considered in their totality, the findings reported here support the conclusion that in East Asian American students, and to a lesser extent in their Non-Hispanic White counterparts, the extent to which academic aspirations predict subsequent postsecondary attainment through the mediation of initial levels of math achievement depends on the degree and direction of bias in self-perceived math ability. In particular, the negative slopes indicate that this mediated effect is strongest at high levels of underconfidence and then progressively fades as self-confidence grows. In fact, in the East Asian American sample, both studies showed that high overconfidence was actually associated with a negative indirect effect, such that educational aspirations significantly predicted the lowest attainment levels. With respect to Mexican American students, in both studies no evidence was found that math self-perception calibration bias influenced the hypothesized construct relations. These findings, together with poorer overall model fit in this group, suggest that other factors omitted from the model may be needed to understand how academic aspirations relate to educational attainment. Perceived barriers (e.g., socioeconomic, institutional, parental, cultural, linguistic) to the pursuit of higher education are likely to be more important than calibration of self-perception in influencing construct relations in Mexican American students (see, for example, Erwin et al., 2021).

How can the findings of the present research inform the work of educators charged with promoting the academic achievement of all students? The educational benefits of academic self-efficacy have been solidly established by a voluminous literature. The results of the two studies reported in this article (as well as other studies reviewed earlier), however, suggest that a more nuanced interpretation of that literature is advisable. It is clear from both studies that in none of the samples was math overconfidence associated with a stronger effect of educational aspirations on postsecondary attainment through initial levels of math ability, and in none of the samples was math underconfidence associated with a weaker effect, which is a finding that is consistent with some of the literature reviewed earlier (e.g., Chiu & Klassen, 2010; Kruger & Dunning, 1999; Talsma et al., 2019). The logical implication is that efforts aimed at improving educational outcomes by raising students' confidence in their math ability may be counterproductive if that confidence is not matched by actual performance. In fact, the findings of both studies suggest that some self-doubt might be helpful. Tables 1 and 5 show that in both studies, compared to East Asian American students, Mexican American participants had significantly lower mean levels of math achievement and postsecondary attainment and they also had significantly higher levels of confidence in their math ability relative to their actual performance. In Non-Hispanic White American participants, the ordering of those means fell between the other two groups. Regardless of racial/ethnic background, an unrealistically optimistic appraisal of one's math ability does not improve math achievement and, as a result, has adverse effects on later educational attainment. By contrast, underestimation of one's math ability has beneficial effects on educational outcomes for East Asian American and Non-Hispanic White American students and no damaging effects for their Mexican American peers. In other words, underconfidence does not seem to benefit all groups equally, but certainly overconfidence helps none of them. The ability to recognize limitations in their math ability may motivate students with high educational aspirations to improve their math performance, which has been found to be a very strong predictor of postsecondary attainment in both of the present studies and in previous research (e.g., Lee et al., 2008; Perna & Titus, 2005; Trusty & Niles, 2003). Furthermore, the spreading grade inflation epidemic (Sanchez & Moore, 2022) is likely to feed students' tendency to overestimate their ability, which may then weaken self-improvement efforts and contribute to poor test scores.

To the extent that some degree of underconfidence promotes educational achievement and attainment, it seems reasonable to raise the question of whether calibration of self-beliefs is malleable and responsive to suitable interventions. There is ample empirical evidence that such is the case (Callender et al., 2016; Dunlosky & Rawson, 2012; Fuchs et al., 2003; Perels et al., 2009, 2005; Schmitz & Perels, 2011). DiGiacomo and Chen (2016), for example, randomized a sample of middle school students from mid-high socioeconomic status to either a control condition or to an intervention designed to improve consistency between self-perceived ability and actual achievement. Students in the intervention group had significantly greater calibration accuracy and better math performance than their untreated peers. These results replicated similar findings with college students by Nietfeld et al. (2006) and by Zimmerman et al. (2011). It should be noted that, in the latter study, 90% of the sample included participants from various minority groups.

5.2. Limitations

Secondary analyses of educational data from large-scale national surveys can make valuable contributions to our understanding of important issues faced by educators and policy makers. Some methodological advantages afforded by the datasets analyzed in this investigation include the availability of longitudinal data, well-executed complex sampling designs, large representative samples that confer substantial statistical power and greater generalizability of the results, and a wealth of information that can support the examination of many research questions. At the same time, these national surveys were not explicitly designed to address the specific hypotheses tested in the present investigation. As a result, some of the constructs that make up the moderated multiple mediator model tested in this research could not be operationalized in the same way across the two datasets. Calibration bias, for example, was measured by comparing self-reported math grades and transcript-based math grades in Study 1, but in Study 2 was measured by comparing students' math self-efficacy and transcript-based math grades. This operationalization discrepancy was necessitated by the fact that the HSLS dataset includes no measure of self-reported grades and the NELS survey lacks a math self-efficacy variable. Although differences in construct operationalization are obviously problematic, the overall similarity in the pattern of results that emerged from the analyses of the two datasets is quite reassuring and is consistent with the findings of the broader self-regulated learning literature reviewed above.

There are, of course, other plausible moderators of construct relations that may interact with participants' racial/ethnic background and that were unexplored in the present research. These omitted variables include, among others, English proficiency level, degree of acculturation, and generational/immigrant status. With respect to calibration of self-beliefs, for example, a self-enhancement tendency that does not reflect actual ability is quite prevalent in individualistic Western societies, but members of East Asian collectivistic cultures are more likely to exhibit a modesty bias despite their high achievement levels (Cai et al., 2011; Chiu & Klassen, 2010; Church, 2009; Eaton & Dembo, 1997; Fu et al., 2016; Klassen, 2004; Kurman, 2002; Leung, 2002). However, there is empirical evidence that following migration to the U.S. or other Western societies, those achievement-promoting cultural assets progressively fade out and this acculturation process may then have negative educational repercussions. First- and second-generation immigrant youth have been found to academically outperform their native peers and their third (or later) generation immigrant peers (Duong et al., 2016; Fuligni, 1997; Han, 2006; Kao & Tienda, 1995). There is also evidence that this "immigrant advantage" is moderated by race/ethnicity, such that the advantage is strongest in Asian samples and weaker or nonexistent in other immigrant groups. A meta-analysis by Duong et al. (2016) corroborated and further elucidated these findings. Regardless of race/ethnicity, both first- and second-generation immigrants outperformed their later generation counterparts; however, the immigrant advantage was greater for second generation than for first generation students, especially in Asian samples. A persuasive explanation offered by the authors is that second-generation immigrant youths have a double advantage: not only do they still retain the achievement-promoting influence of *protective cultural values* (e.g., the modesty bias), but they also have access to more achievement-promoting resources than their first-generation peers (e.g., greater stability and financial security, greater English proficiency).

In the analyses of the two datasets, several group comparisons were conducted and this may have increased the likelihood of Type I error. Although a Bonferroni correction could have been used to reduce this false-positive risk, there is evidence that unequal group sizes, a common occurrence in cross-cultural research, tend to inflate Type II error rates and reduce the ability to detect group differences (Aguinis, 1995; Chen, 2007; Kaplan & George, 1995; Yoon & Lai, 2018). Laczo et al. (2005), for example, found that the standard error of the standardized mean difference between two groups grows progressively as group sizes move away from a 50–50 split and, at a 95–5 split, the standard error is 130% larger compared to the 50–50 split. This was certainly an issue in the present investigation; the split between the largest and smallest groups was 96–4 in the NELS dataset and 95–5 in the HSLS dataset. Thus, α-level adjustment would have been an overly conservative course of action.

5.3. Future directions

In none of the three racial/ethnic groups and in neither of the two studies reported here was there evidence that the important relation between student aspirations and postsecondary attainment through the mediation of math growth (i.e., the slope factor) was moderated by calibration bias. This is a very interesting finding, deserving of future research attention. It suggests that, whatever beneficial or damaging effects calibration may have on that relation, those effects tend to occur early in the process and are not modified by subsequent changes in math performance. A possible mechanism through which this could happen is the practice of tracking students into different math ability groups or different types of math classes. Once calibration has exerted its beneficial or harmful influence on the relation between aspirations and initial math achievement and once students are segregated based on these initial math achievement levels, they may progress at roughly the same rate within each ability group. Under this scenario, calibration might not influence math achievement slope but its early effects on math achievement would then explain differences in distal educational outcomes.

In the present investigation, analyses of both datasets indicated that overall model fit and construct relations were substantially weaker in Mexican American students. Model misspecification in this group may have resulted from omission of variables (e.g., sociocultural factors, parental values and beliefs, generational/immigrant status, home environment) that could operate as important predictors, mediators, or moderators. Consider, for example, the extensive literature on the educational effects of familism (*familismo*), a core value in Hispanic/Latinx culture that is characterized by a strong family orientation (loyalty, interdependence, closeness, obligation, mutual respect and support). Although some studies have found a positive association between familism and academic achievement (Niemeyer et al., 2009; Roche et al., 2012), other studies have reported negative findings (Chun & Devall, 2019; Rodríguez, 2002). Inconsistency in the literature may be due to differences in the operationalization of familism (i.e., which particular facets of this multidimensional construct are assessed), but also might be attributed to failure to consider moderators of that relation. In a recent large-scale meta-analysis by Cahill et al. (2021), both students' and parents' nativity status (U.S. born vs. non-U.S. born) were found to be significant moderators, but in opposite directions. The effect of familism on educational outcomes was positive and strongest when samples included a large proportion of parents born outside the U.S. and became increasingly negative as the proportion of U.S. born parents increased. Conversely, when the moderating influence of students' nativity was considered, the familism-achievement relation was strongest and positive when the proportion of students born outside the U.S. decreased. In view of these findings, and similar ones (e.g., Gordon, 2017; Stein et al., 2014; Toyokawa & Toyokawa, 2019; Valenzuela & Dornbusch, 1994), it seems reasonable to speculate that in the Mexican American sample the moderating influence of calibration bias may itself be a function of nativity status or other variables omitted from the model.

In both studies reported here, compared to the other two groups, Mexican American students had the lowest math achievement levels but the highest confidence in their math ability relative to their actual performance. Another promising line of inquiry would focus on the psychological mechanisms that might explain this paradox. Positive stereotypes (e.g., "Asians are good at math") and negative stereotypes (e.g., "Latinx are bad at math") have been found to influence task performance in a direction that is consistent with the stereotype (e.g., Appel et al., 2015; Armenta, 2010; Seo & Lee, 2021; Shih et al., 2002, 1999; Spencer et al., 2016; Steele & Aronson, 1995). Armenta (2010), for example, confirmed the stereotype-boosting and stereotype-depressing effects on math performance in Asian American and Mexican American student samples, respectively. Interestingly, in both groups, stereotype susceptibility was enhanced by strong group identification.

The damaging cognitive effects of stigmatization (e.g., stereotype threat) may help explain Latinx students' poorer math performance, but cannot explain how/why, in this group, self-perceived math ability gets a boost in the face of poor performance. Two longstanding hypotheses that have received considerable empirical support and that might assist in understanding this inconsistency are the academic disidentification (i.e., selective devaluation) hypothesis (e.g., Major et al., 1998; Osbourne, 1995; Woodcock et al.,

2012) and the external attribution hypothesis (e.g., Crocker & Major, 1989; van Laar, 2000). The first posits that poorly performing minority groups may protect their positive self-beliefs by becoming progressively disengaged from academics and devaluing both schoolwork and school performance. The second hypothesis suggests that the same protective function can come from attributing poor performance to external circumstances (e.g., sociocultural barriers such as discrimination, marginalization, and implicit bias). Although both hypotheses can explain the maintenance of high confidence despite poor achievement in Mexican American students (and other negatively stereotyped minorities), a recent study by Seo et al. (2019) found little evidence for those theoretical formulations in a nationally representative sample of nearly 2220 Latinx students who participated in the Education Longitudinal Study. Clearly, this is a line of research that requires greater attention.

Despite a great deal of progress in identifying context-invariant predictors of academic achievement and educational attainment across countries and cultural groups, stubborn gaps in educational outcomes endure across racial/ethnic groups and the equity goal is not within reach. What is required is a clearer understanding of culture-specific mechanisms that are likely to operate in the educational process. The findings of the work reported here indicate that calibration bias may have differential moderating effects across the three racial/ethnic groups investigated and, differently from the other two groups, does not appear to play a significant role in explaining the relation between Mexican American students' aspirations and their educational outcomes. Future work aimed at disentangling these intricate connections by examining the contribution of culture-specific influences will hopefully point to productive avenues for reducing group discrepancies in educational achievement and attainment.

Declaration of Competing Interest

None.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21*(6), 1141–1158. https://doi.org/10.1177/014920639502100607.

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology, 6*, 1–15. https://doi.org/10.3389/fpsyg.2015.00900.

Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16*(1), 94–98. https://doi.org/10.1037/a0017564.

Baker, B. D. (2021). *Educational inequality and school finance: Why money matters for America's students*. Harvard Education Press.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40*(3), 373–400. https://doi.org/10.1207/s15327906mbr4003_5.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11*(2), 142–163. https://doi.org/10.1037/1082-989X.11.2.142.supp (Supplemental).

Beauducel, A., & Herzberg, P. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78–117. https://doi.org/10.1177/0049124187016001004.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education, 69*(2), 133–151. https://doi.org/10.1080/00220970109600653.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290.

Cahill, K. M., Updegraff, K. A., Causadias, J. M., & Korous, K. M. (2021). Familism values and adjustment among Hispanic/Latino individuals: A systematic review and meta-analysis. *Psychological Bulletin, 147*(9), 947–985. https://doi.org/10.1037/bul0000336.supp (Supplemental).

Cai, H., Sedikides, C., Gaertner, L., Wang, C., Carvallo, M., Xu, Y., … Jackson, L. E. (2011). Tactical self-enhancement in China: Is modesty at the service of self-enhancement in East Asian culture? *Social Psychological and Personality Science, 2*, 59–64. https://doi.org/10.1177/1948550610376599.

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning, 11*(2), 215–235. https://doi.org/10.1007/s11409-015-9142-6.

Carnegie Foundation for the Advancement of Teaching. (2011). *The Carnegie classification of institutions of higher education* (2010 ed.) Author.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421–483. https://doi.org/10.1177/109442819814004.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834.

Chen, P.-Y., Wu, W., Garnier-Villarreal, M., Kite, B. A., & Jia, F. (2020). Testing measurement invariance with ordinal missing data: A comparison of estimators and missing data techniques. *Multivariate Behavioral Research, 55*(1), 87–101. https://doi.org/10.1080/00273171.2019.1608799.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2–17. https://doi.org/10.1016/j.learninstruc.2008.11.002.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44*, 347–357. https://doi.org/10.1111/j.2044-8317.1991.tb00966.x.

Chun, H., & Devall, E. (2019). A parental involvement and academic socialization model: A cultural approach. *School Psychology, 34*(5), 555–565. https://doi.org/10.1037/spq0000330.

Church, A. T. (2009). Prospects for an integrated trait and cultural psychology. *European Journal of Personality, 23*(3), 153–182. https://doi.org/10.1002/per.700.

Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative approached to structural modeling of ordinal data: A Monte Carlo study. *Structural Equation Modeling, 4*, 261–282.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608–630. https://doi.org/10.1037/0033-295X.96.4.608.

Cruz, C. (2021). From digital disparity to educational excellence: Closing the opportunity and achievement gaps for low-income, Black, and Latinx students. *Harvard Latino Law Review, 24*, 33–64.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*, 526–544. https://doi.org/10.1037/0021-843X.112.4.526.

Davis-Kean, P. E., Domina, T., Kuhfeld, M., Ellis, A., & Gershoff, E. T. (2022). It matters how you start: Early numeracy mastery predicts high school math course-taking and college attendance. *Infant and Child Development, 31*(2), Article e2281. https://doi.org/10.1002/icd.2281.

DiGiacomo, G., & Chen, P. P. (2016). Enhancing self-regulatory skills through an intervention embedded in a middle school mathematics curriculum. *Psychology in the Schools, 53*, 601–616. https://doi.org/10.1002/pits.21929.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*, 327–346. https://doi.org/10.1207/S15328007SEM0903_2.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003.

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. M. Olson, & M. P. Zanna (Eds.)*, 44*. *Advances in experimental social psychology* (pp. 247–296). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00005-6.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235.

Duong, M. T., Badaly, D., Liu, F. F., Schwartz, D., & McCarty, C. A. (2016). Generational differences in academic achievement among immigrant youths: A meta-analytic review. *Review of Educational Research, 86*(1), 3–41. https://doi.org/10.3102/0034654315577680.

Eaton, M. J., & Dembo, M. H. (1997). Differences in the motivational beliefs of Asian American and non-Asian students. *Journal of Educational Psychology, 89*(3), 433–440. https://doi.org/10.1037/0022-0663.89.3.433.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*(1), 1–22. https://doi.org/10.1037/1082-989X.12.1.1.supp (Supplemental).

Eekhout, I., Twisk, J. W. R., Heymans, M. W., de Vet, H. C. W., de Boer, M. R., & Enders, C. K. (2015). Analyzing incomplete item scores in longitudinal data by including item score information as auxiliary variables. *Structural Equation Modeling, 22*(4), 588–602. https://doi.org/10.1080/10705511.2014.937670.

Ehrlinger, J., Johnson, K. L., Banner, M., Dunning, D. A., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002.

Erwin, B., Thomsen, J., & Education Commission of the States. (2021). *Addressing inequities in higher education*. Education Commission of the States: Policy guide.

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*(2), 87–99. https://doi.org/10.1007/s11121-008-0109-6.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Information Age Publishing.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal, 91*(1), 29–42. https://doi.org/10.1353/hsj.2007.0022.

Freeman, E. E., Karayanidis, F., & Chalmers, K. A. (2017). Metacognitive monitoring of working memory performance and its relationship to academic achievement in grade 4 children. *Learning and Individual Differences, 57*, 58–64. https://doi.org/10.1016/j.lindif.2017.06.003.

Fu, G., Heyman, G. D., Cameron, C. A., & Lee, K. (2016). Learning to be unsung heroes: Development of reputation management in two cultures. *Child Development, 87*(3), 689–699. https://doi.org/10.1111/cdev.12494.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003). Enhancing third-grade student' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*(2), 306–315. https://doi.org/10.1037/0022-0663.95.2.306.

Fuligni, A. J. (1997). The academic achievement of adolescents from immigrant families: The roles of family background, attitudes, and behavior. *Child Development, 68*(2), 351–363.

Gambini, A., Desimoni, M., & Ferretti, F. (2022). Predictive tools for university performance: An explorative study. *International Journal of Mathematical Education in Science and Technology*, 1–27. Advance online publication https://doi.org/10.1080/0020739x.2021.2022794.

Gordon, M. (2017). Community disadvantage and academic achievement among Hispanic adolescents: The role of familism. *Journal of Child and Family Studies, 26*(12), 3303–3311. https://doi.org/10.1007/s10826-017-0845-y.

Gottfredson, N. C., Panter, A. T., Daye, C. E., Allen, W. F., & Wightman, L. F. (2009). The effects of educational diversity in a national sample of law students: Fitting multilevel latent variable models in data with categorical indicators. *Multivariate Behavioral Research, 44*(3), 305–331. https://doi.org/10.1080/00273170902794971.

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4*(2), 108–120.

Guglielmi, R. S. (2008). Native language proficiency, English literacy, academic achievement, and occupational attainment in Limited English Proficient students: A latent growth modeling perspective. *Journal of Educational Psychology, 100*, 322–342. https://doi.org/10.1037/0022-0663.100.2.322.

Guglielmi, R. S. (2012). Math and science achievement in English Language Learners: Multivariate latent growth modeling of predictors, mediators, and moderators. *Journal of Educational Psychology, 104*, 580–602. https://doi.org/10.1037/a0027378.

Guglielmi, R. S., & Brekke, N. (2018). A latent growth moderated mediation model of math achievement and postsecondary attainment: Focusing on context-invariant predictors. *Journal of Educational Psychology, 110*, 683–708. https://doi.org/10.1037/edu0000238.

Guglielmi, R. S., & Brekke, N. C. (2017). A framework for understanding cross-national and cross-ethnic gaps in math and science achievement: The case of the U.S. *Comparative Education Review, 61*(1), 176–213.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*(2), 101–121. https://doi.org/10.1007/s11409-008-9021-5.

Hafner, A., Ingels, S. J., Schneider, B., & Stevenson, D. L. (1990). *A Profile of the American eighth grader: NELS:88 student descriptive summary. NCES 90–458*. NCES.

Han, W.-J. (2006). Academic achievements of children in immigrant families. *Educational Research Review, 1*(8), 286–318.

Harwell, M., Maeda, Y., Bishop, K., & Xie, A. (2017). The surprisingly modest relationship between SES and educational achievement. *Journal of Experimental Education, 85*(2), 197–214. https://doi.org/10.1080/00220973.2015.1123668.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research, 50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683.

Hayes, A. F. (2018a). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs, 85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100.

Hayes, A. F. (2018b). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41*(3), 924–936. https://doi.org/10.3758/BRM.41.3.924.

Hayes, A. F., & Preacher, K. J. (2013). Conditional process modeling: Using structural equation modeling to examine contingent causal processes. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 219–266). Information Age Publishing.

Heath, A., Rothon, C., & Kilpi, E. (2008). The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology, 34*(1), 211–235. https://doi.org/10.1146/annurev.soc.34.040507.134728.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17*, 63–84. https://doi.org/10.1016/j.edurev.2015.11.002.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., … Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSLS:09). In *Base-Year Data File Documentation (NCES 2011–328)*. U.S. Department of Education.

Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., … Leinwand, S. (2013). *High School Longitudinal Study of 2009 (HSLS:09) Base Year to First Follow-Up Data File Documentation (NCES 2014–361)*. U.S. Department of Education.

Jackson, M. (2012). Bold choices: How ethnic inequalities in educational attainment are suppressed. *Oxford Review of Education, 38*(2), 189–208. https://doi.org/10.1080/03054985.2012.676249.

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods, 51*(5), 2337–2355. https://doi.org/10.3758/s13428-018-1187-4.

Kao, G., & Tienda, M. (1995). Optimism and achievement: the educational performance of immigrant youth. *Social Science Quarterly, 76*, 1–19.

Kaplan, D., & George, R. (1995). A Study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling, 3*(2), 101–118.

Kendi, I. X. (2019). *How to be an antiracist*. One World.

Kim, S. W., Cho, H., & Kim, L. Y. (2019). Socioeconomic status and academic outcomes in developing countries: A meta-analysis. *Review of Educational Research, 89*(6), 875–916.

Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective. *International Journal of Psychology, 39*(3), 205–230. https://doi.org/10.1080/00207590344000330.

Koziol, N. A., & Bovaird, J. A. (2018). The impact of model parameterization and estimation methods on tests of measurement invariance with ordered polytomous data. *Educational and Psychological Measurement, 78*(2), 272–296. https://doi.org/10.1177/0013164416683754.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A metaanalysis and review of the literature. *Review of Educational Research, 75*, 63–82. https://doi.org/10.3102/00346543075001063.

Kurman, J. (2002). Measured cross-cultural differences in self-enhancement and the sensitivity of the self-enhancement measure to modesty response. *Cross-Cultural Research, 36*, 73–95. https://doi.org/10.1177/1069397102036001004.

van Laar, C. (2000). The paradox of low academic achievement but high self-esteem in African American students: An attributional account. *Educational Psychology Review, 12*(1), 33–61. https://doi.org/10.1023/A:1009032900261.

Laczo, R. M., Sackett, P. R., Bobko, P., & Cortina, J. M. (2005). A comment on sampling error in the standardized mean difference with unequal sample sizes: Avoiding potential errors in meta-analytic and primary research. *Journal of Applied Psychology, 90*(4), 758–764. https://doi.org/10.1037/0021-9010.90.4.758.

Lang, K. M., & Little, T. D. (2018). Principled missing data treatments. *Prevention Science, 19*(3), 284–294. https://doi.org/10.1007/s11121-016-0644-5.

Lee, S. M., Daniels, M. H., Puig, A., Newgent, R. A., & Nam, S. K. (2008). A data-based model to predict postsecondary educational attainment of low-socioeconomic-status students. *Professional School Counseling, 11*(5), 306–316. https://doi.org/10.5330/PSC.n.2010-11.306.

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity, 43*, 495–507. https://doi.org/10.1007/s11135-007-9133-z.

Lei, P.-W., & Shiverdecker, L. K. (2020). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. *Structural Equation Modeling, 27*(4), 584–601. https://doi.org/10.1080/10705511.2019.1680292.

Leung, F. K. S. (2002). Behind the high achievement of East Asian students. *Educational Research and Evaluation, 8*(1), 87–108. https://doi.org/10.1076/edre.8.1.87.6920.

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369–387. https://doi.org/10.1037/met0000093.supp.

Liu, J., Peng, P., & Luo, L. (2020). The relation between family socioeconomic status and academic achievement in China: A meta-analysis. *Educational Psychology Review, 32*(1), 49–76. https://doi.org/10.1007/s10648-019-09494-0.

Liu, Y., & Thompson, M. S. (2021). General factor mean difference estimation in bifactor models with ordinal data. *Structural Equation Modeling, 28*(3), 423–439. https://doi.org/10.1080/10705511.2020.1833732.

LoGerfo, L., Christopher, E. M., & Flanagan, K. D. (2011). *High School Longitudinal Study of 2009 (HSLS:09). A first look at Fall 2009 ninth-graders' parents, teachers, school counselors, and school administrators (NCES 2011–355)*. U.S. Department of Education, National Center for Education Statistics. Government Printing Office.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*(1), 34–50. https://doi.org/10.1177/0146167298241003.

Mayor, C., & Suarez, E. B. (2019). A scoping review of the demographic and contextual factors in Canada's educational opportunity gaps. *Canadian Journal of Education, 42*(1), 42–87.

McCabe, C. J., Kim, D. S., & King, K. M. (2018). Improving present practices in the visual display of interactions. *Advances in Methods and Practices in Psychological Science, 1*(2), 147–165. https://doi.org/10.1177/2515245917746792.

Mihalca, L. (2014). Why prompting metacognition in computer-based learning environments. *Cognition, Brain, & Behavior, 18*(4), 299–314.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 502–506. https://doi.org/10.1037/a0021802.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* ((8th ed.).). Author.

Niemeyer, A. E., Wong, M. M., & Westerhaus, K. J. (2009). Parental involvement, familismo, and academic performance in Hispanic and Caucasian adolescents. *North American Journal of Psychology, 11*(3), 613–632.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*, 159–179. https://doi.org/10.1007/s10409-006-9595-6.

Nomi, T., Raudenbush, S. W., & Smith, J. J. (2021). Effects of double-dose algebra on college persistence and degree attainment. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 118*(27). https://doi.org/10.1073/pnas.2019030118.

Organization for Economic Cooperation and Development. (2019). *PISA 2018 results: Combined executive summaries* (Vol. I, II & III). OECD Publishing.

Osbourne, J. W. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin, 21*(5), 449–455. https://doi.org/10.1177/0146167295215003.

Osterhage, J. L., Usher, E. L., Douin, T. A., & Bailey, W. M. (2019). Opportunities for self-evaluation increase student calibration in an introductory biology course. *CBE Life Sciences Education, 18*(2), 1–10.

Padgett, R. N., & Morgan, G. B. (2021). Multilevel CFA with ordered categorical data: A simulation study comparing fit indices across robust estimation methods. *Structural Equation Modeling, 28*(1), 51–68. https://doi.org/10.1080/10705511.2020.1759426.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*(2), 124–139. https://doi.org/10.1006/ceps.1998.0991.

Perels, F., Dignath, C., & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), 17–31. https://doi.org/10.1007/BF03173472.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15*(2), 123–139. https://doi.org/10.1016/j.learninstruc.2005.04.010.

Perna, L., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education, 76*, 485–518. https://doi.org/10.1080/00221546.2005.11772296.

Poulsen, S. (2019). The effect of additional math in high school on college success. *The Mathematics Educator, 28*(2), 76–90.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*(4), 437–448. https://doi.org/10.3102/10769986031004437.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*(1), 185–227. https://doi.org/10.1080/00273170701341316.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Sage.

Quinn, D. M. (2020). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher, 49*(7), 482–492. https://doi.org/10.3102/0013189X20932469.

Raykov, T., & West, B. T. (2016). On enhancing plausibility of the missing at random assumption in incomplete data analyses via evaluation of response-auxiliary variable correlations. *Structural Equation Modeling, 23*(1), 45–53. https://doi.org/10.1080/10705511.2014.937848.

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315.supp (Supplemental).

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.

Rioux, C., & Little, T. D. (2021). Missing data treatments in intervention studies: What was, what is, and what should be. *International Journal of Behavioral Development, 45*(1), 51–58.

Roche, K. M., Ghazarian, S. R., & Fernandez-Esquer, M. E. (2012). Unpacking acculturation: Cultural orientations and educational attainment among Mexican-origin youth. *Journal of Youth and Adolescence, 41*(7), 920–931. https://doi.org/10.1007/s10964-011-9725-8.

Rock, D. A., & Pollack, J. M. (1995). *Psychometric report for the NELS:88 base year through second follow-up* (NCES 95–382). National Center for Education Statistics

Rodríguez, J. L. (2002). Family environment and achievement among three generations of Mexican American high school students. *Applied Developmental Science, 6*(2), 88–94. https://doi.org/10.1207/S1532480XADS0602_4.

Rothon, C. (2007). Can achievement differentials be explained by social class alone? An examination of minority ethnic educational performance in England and Wales at the end of compulsory schooling. *Ethnicities, 7*(3), 306–322. https://doi.org/10.1177/1468796807080231.

Sanchez, E. I., & Moore, R. (2022). Grade inflation continues to grow in the past decade. ACT Research Report. https://www.act.org/content/act/en/research/pdfs/R2134-Grade-Inflation-Continues-to-Grow-in-the-Past-Decade-Final-Accessible.html.

Schmitz, B., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning, 6*(3), 255–273. https://doi.org/10.1007/s11409-011-9076-6.

Schneider, B. (2003). Strategies for success: High school and beyond. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 55–94). Brookings Institution Press.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565–600. https://doi.org/10.1037/bul0000098.supp (Supplemental).

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3.

Schwartz, J. A., & Beaver, K. M. (2015). Making (up) the grade? Estimating the genetic and environmental influences of discrepancies between self-reported grades and official GPA scores. *Journal of Youth and Adolescence, 44*, 1125–1138. https://doi.org/10.1007/s10964-014-0185-9.

Sciarra, D. T., & Whitson, M. L. (2007). Predictive factors in postsecondary educational attainment among Latinos. *Professional School Counseling, 10*(3), 307–316.

Seo, E., & Lee, Y. (2021). Stereotype threat in high school classrooms: How it links to teacher mindset climate, mathematics anxiety, and achievement. *Journal of Youth and Adolescence, 50*(7), 1410–1423. https://doi.org/10.1007/s10964-021-01435-x.

Seo, E., Shen, Y., & Benner, A. D. (2019). The paradox of positive self-concept and low achievement among Black and Latinx youth: A test of psychological explanations. *Contemporary Educational Psychology, 59*, 1–37. https://doi.org/10.1016/j.cedpsych.2019.101796.

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition, 44*(7), 1127–1137. https://doi.org/10.3758/s13421-016-0624-9.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*(3), 638–647. https://doi.org/10.1037/0022-3514.83.3.638.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 81–84.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415–437. https://doi.org/10.1146/annurev-psych-073115-103235.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research, 50*(2), 277–288. https://doi.org/10.1509/jmr.12.0420.

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences, 55*(7), 727–732. https://doi.org/10.1016/j.paid.2013.07.006.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Stein, G. L., Cupito, A. M., Mendez, J. L., Prandoni, J., Huq, N., & Westerberg, D. (2014). Familism through a developmental lens. *Journal of Latina/o Psychology, 2*(4), 224–250. https://doi.org/10.1037/lat0000025.

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling, 22*(4), 568–580. https://doi.org/10.1080/10705511.2014.937669.

Talsma, K., Schüz, B., & Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy ≠ self-fulfilling prophecy. *Learning and Individual Differences, 69*, 182–195. https://doi.org/10.1016/j.lindif.2018.11.002.

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015.

Thompson, M. S., Gorin, J. S., Obeidat, K., & Chen, Y. (2006). Understanding differences in postsecondary educational attainment: A comparison of predictive measures for Black and White students. *Journal of Negro Education, 75*(3), 546–562.

Toyokawa, N., & Toyokawa, T. (2019). Interaction effect of familism and socioeconomic status on academic outcomes of adolescent children of Latino immigrant families. *Journal of Adolescence, 71*, 138–149. https://doi.org/10.1016/j.adolescence.2018.10.005.

Trusty, J., & Niles, S. G. (2003). High-school math courses and completion of the bachelor's degree. *Professional School Counseling, 7*(2), 99–107.

U.S. Census Bureau. (2021). *Educational attainment in the United States: 2021*. Current Population Survey, 2021, Annual Social and Economic Supplement (CPS ASEC). https://www.census.gov/data/tables/2021/demo/educational-attainment/cps-detailed-tables.html.

U.S. Department of Education. (2020). National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990–2019 Mathematics Assessments. https://www.nationsreportcard.gov/mathematics/nation/groups/?grade=4.

Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Muraszewski, A. K. (2018). Can reading self-efficacy be modified? A meta-analysis of the impact of interventions on reading self-efficacy. *Review of Educational Research, 88*(2), 167–204. https://doi.org/10.3102/0034654317743199.

Valenzuela, A., & Dornbusch, S. M. (1994). Familism and social capital in the academic achievement of Mexican origin and Anglo adolescents. *Social Science Quarterly, 75*(1), 18–36.

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. W. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. *Journal of Personality and Social Psychology, 103*(4), 635–646. https://doi.org/10.1037/a0029120.

Xu, Z., Backes, B., Oliveira, A., & Goldhaber, D. (2020). *Targeted interventions in high school: Preparing students for college. (CALDER Working Paper No. 232–0220)*. American Institutes for Research. https://eric.ed.gov/?id=ED603894.

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling, 25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859.

Yuan, K.-H., & Savalei, V. (2014). Consistency, bias and efficiency of the normal-distribution-based MLE: The role of auxiliary variables. *Journal of Multivariate Analysis, 124*, 353–370. https://doi.org/10.1016/j.jmva.2013.11.006.

Yuan, K.-H., & Zhang, Z. (2012). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika, 77*, 803–826.

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling, 53*(1), 141–160.

Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2017). The role of self-efficacy and related beliefs in self-regulation of learning and performance. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 313–333). The Guilford Press.