Is There any Evidence of Historical Changes in Gender Differences in American High School Students' Math Competence-Related Beliefs from the 1980s to the 2010s?

Charlott Rubach[1], Glona Lee[1], Christine R. Starr[1], Yannan Gao[1], Nayssan Safavian[1], Anna-Lena Dicke[1], Jacquelynne S. Eccles[1], & Sandra D. Simpkins[1]

[1]School of Education, University of California, 401 E. Peltason Drive, 3200 Education Bldg., Irvine, CA 92697, USA; phone: (949) 824-8092, fax: (949) 824-9103, crubach@uci.edu, glonal@uci.edu, cstarr1@uci.edu, yannang@uci.edu, nayssan.safavian@uci.edu, adicke@uci.edu, jseccles@uci.edu, simpkins@uci.edu

Author Note

Charlott Rubach	https://orcid.org/0000-0003-0451-6429
Glona Lee	https://orcid.org/0000-0003-4585-3456
Christine R. Starr	https://orcid.org/0000-0002-8662-0387
Yannan Gao	https://orcid.org/0000-0003-3949-5488
Nayssan Safavian	https://orcid.org/0000-0001-7765-2363
Anna-Lena Dicke	https://orcid.org/0000-0001-8816-455X
Jacquelynne S. Eccles	https://orcid.org/0000-0002-6405-9330
Sandra D. Simpkins	https://orcid.org/0000-0002-6053-4827

We have no conflict of interest to disclose. This work was supported by two National Science Foundation grants (EHR-1760757 & EHR-2054956) to Sandra Simpkins & Jacquelynne Eccles. Additionally, multiple grants supported collecting and sustaining three of the datasets (CAB, MADICS, and MSALT) that this paper employs: two NSF grants to Jacquelynne Eccles (BNS 85-10504) and Jacquelynne Eccles and Pamela Davis-Kean (0089972), two grants from the National Institute for Child Health and Human Development to Jacquelynne Eccles (HD17296) and Jacquelynne Eccles, Allan Wigfield, Phyllis Blumenfeld, and Rena Harold (HD17553), two grants from the MacArthur Network, a grant from the National Institute for Mental Health (MH31724), and a grant from the W.T. Grant Foundation, each awarded to Jacquelynne Eccles. Finally, two National Science Foundation grants also supported the collecting and sustaining of the CAMP dataset to Jacquelynne S. Eccles, Nayssan Safavian, and Stuart Karabenick (HRD-1535273), and Martin Maehr and Stuart Karabenick (DRL-0335369). We thank the principals, teachers, students, and parents of the cooperating school districts for their participation in these projects.

Please direct correspondence to the corresponding author: Charlott Rubach, School of Education, University of California, 401 E. Peltason Drive, 3200 Education Bldg., Irvine, CA 92697, United States of America; crubach@uci.edu

Abstract. In this replication study, we examined gender differences in students' math competence-related beliefs from 9th to 12th grade and tested gender differences within four racial/ethnic groups. In order to test the potential historical changes in these patterns and to counteract the replication crisis in psychology, this study employed six U.S. datasets collected from the 1980s to 2010s. Using a total sample of 24,290 students (49.5% male students; 11% African-, 9% Asian-; 30% Latinx- and 50% European-Americans), we found gender differences in students' math competence-related beliefs favoring boys at all grade levels. By comparing effect sizes across datasets, we found no evidence that these gender differences varied by dataset or by historical time. The results across race/ethnicity with a subsample of 23,070 students indicated meaningful gender differences in students' math competence-related beliefs favoring boys at all grade levels among Asian-, European-, and Latinx-Americans, but not among African-Americans where differences favored girls in 12th grade. Overall, our findings provide no evidence of historical changes concerning gender differences in students' math competence-related beliefs across datasets. Our findings illustrate the importance of replicating empirical findings across datasets and using an intersectional lens to investigate math motivation.

Keywords. Math self-concept, math self-efficacy, replication, intersectionality, gender, race/ethnicity

Highlights

- We conducted a conceptual replication of gender differences in students' math competence-related beliefs using six U.S. datasets from the 1980s to 2010s.
- Gender differences in students' math competence-related beliefs favoring boys replicated from 9^{th} to 12^{th} grade.
- Gender differences favoring boys emerged for Asian-, European- and Latinx-Americans but not for African-Americans.
- Gender differences did not vary by the year of data collection and, thus, displayed no historical changes in the gendered patterns.

Is There any Evidence of Historical Changes in Gender Differences in American High School Students' Math Competence-Related Beliefs from the 1980s to the 2010s?

INTRODUCTION

Women, African-Americans, and Latinx-Americans continue to be underrepresented in many math-intensive STEM occupations (Honey et al., 2020; NSF, 2021). Scholars have argued that these disparities are partly due to differences in contextual influences, including limited opportunities, structural barriers, and discrimination that diminish individuals' motivation to pursue STEM (Cheryan et al., 2020; Wang & Degol, 2017). Though scholars have charted trends in STEM occupations over time and countless initiatives have been funded to bolster marginalized students' motivation to pursue STEM (National Center for Science and Engineering Statistics, 2021), little empirical evidence exists on whether the differences in students' motivational beliefs have shifted over time.

To address this gap, the goal of this study was to test if gender differences in high school students' math competence-related beliefs differed across six U.S. datasets than span from the 1980s to 2010s. This study focused on math competence-related beliefs in 9th through 12th grade as high school is a pivotal period for STEM motivation development and students' high school math competence-related beliefs are central determinants of students' future STEM aspirations and choices (Botella et al., 2019; Lazarides et al., 2021). Moreover, much of the existing literature focuses on gender differences across all youth. However, emerging work suggests that gender differences may vary by race/ethnicity (Else-Quest et al., 2013; Seo et al., 2019). We extend prior research by examining gender differences overall and within the four largest racial/ethnic groups in the U.S. (i.e., African-, Asian-, European-, and Latinx-Americans), and by systematically testing if these gender differences replicate across the six datasets.

Motivational theories

Several motivational theories highlight the importance of students' perceptions of their abilities and skills, which we label competence-related beliefs (for review, see Muenks et al., 2018). Competence-related beliefs include students' (a) expectancies of success as described in situated expectancy-value theory (Eccles & Wigfield, 2020), (b) self-efficacy from social cognitive theory (Bandura, 1997), (c) ability self-concepts from self-concept theory (Marsh & Martin, 2011) and (d) competence experiences from self-determination theory (Deci & Ryan, 1985). Though these constructs are somewhat different based on, for example, their normativity as well as the specificity of students' evaluations based on context and time (Bong & Skaalvik, 2003), they share conceptual similarities in their focus on perceived competence explained by psychological processes, such as social comparisons or mastery experiences, and in their domain-specificity (Bong & Skaalvik, 2003; Muenks et al., 2018).

According to these theories, motivational processes are situated within and influenced by the contexts in which students are embedded (Bandura, 1997; Eccles & Wigfield, 2020). One central aspect of students' context is the historical moment – both in terms of students' age and the historical time in which students grow up. Several scholars have described the changes in students' competence-related beliefs due to students' age (Fredricks & Eccles, 2002; Scherrer & Preckel, 2019; Wigfield, Eccles, Fredericks, Simpkins, Roeser & Schiefele, 2015); yet, few have examined differences based on historical time (Parker et al., 2018; Scherrer & Preckel, 2019). One meta-analytic study suggests that students' math competence-related beliefs decline at a similar rate across data from different decades (Scherrer & Preckel, 2019). In addition, Parker and colleagues (2018) found no evidence of historical changes concerning gender differences in Australian students' STEM competence-beliefs from 1981 to 1993. However, to our knowledge, no replication study has addressed historical changes in gender differences in U.S. students' math competence-related beliefs and whether gender differences replicate across racial/ethnic groups, which is paramount given the growing diversity in the U.S.

Gender differences overall

Multiple theories, such as social comparison theory, social cognitive theory, and situated expectancy-value theory, argue that the social environment and social processes determine students' competence-related beliefs (see Bandura, 1997; Eccles & Wigfield, 2020; Skaalvik & Rankin, 1990). For example, situated expectancy-value theory (Eccles & Wigfield, 2020) helps explain the interrelations of various social and demographic components with individuals' competence-related beliefs. Specifically, this theory argues that motivational processes are situated within a student's sociocultural context which is known as the cultural milieu. Aspects of the cultural milieu, including societal norms and stereotypes about demographic groups (e.g., gender, race/ethnicity, socioeconomic status), influence their competence-related beliefs (Eccles & Wigfield, 2020).

In this study, we primarily focus on gender as the primary explanatory factor of students' math competence-related beliefs as many STEM fields have been historically defined as gendered, where male students hold more privilege than female students. For example, many STEM fields, including math, are stereotyped as masculine domains (Miller et al., 2018). Girls perceive their competence in math activities as lower than boys in spite of demonstrating comparable academic achievement – illustrating the power of society's beliefs about social categories (Bandura, 1997). In line with theoretical assumptions, studies utilizing datasets from the 1980s to 2000s found gender differences concerning math competence-related beliefs than their female peers. Specifically, cross-sectional findings suggest that high school boys' math competence-related beliefs are higher than girls' beliefs (Else-Quest et al., 2013, 2010; Skaalvik & Rankin, 1994; Marsh et al., 2021; Watt et al., 2012). Parallel patterns favoring boys have emerged in most longitudinal studies (Fredricks & Eccles, 2002; Graham & Morales-Chicas, 2015; Jiang et al., 2020; Nagy et al., 2010; Umarji et al., 2021; Watt, 2004). Also, a meta-analysis of 176 studies demonstrated gender differences favoring boys in students' math expectancies for success (Parker et al., 2020). The effect sizes typically range from

$|.15 \geq d \leq .37|$ (Else-Quest et al., 2010, 2013; Parker et al., 2020). Collectively, these works suggest that gender differences have persisted over time and are unlikely to vary across historical time.

However, numerous efforts have been launched to counteract gender disparities in STEM. For example, there has been an increasing interest in promoting diversity in STEM education across U.S. colleges and high schools (Granovskiy, 2018). General policies, however, often focus on counteracting gendered achievement and participation gaps in STEM. Efforts at the college level often focused on increasing the enrollment of women in STEM college majors, improving recruitment and retention of women, and counteracting the achievement gap between male and female students in STEM classes (see NASEM, 2020; NSF, 2021). Though women's enrollment in many STEM college majors (e.g., computer sciences, engineering, mathematics, and statistics) has increased over the last few decades (NCSES, 2021), men continue to account for the majority of majors in many STEM fields.

At the high school level, interventions have been launched to counteract gender disparities in STEM (for reviews of these interventions, see Lee et al., 2021; Levene & Pantoja, 2021). One example is the growth mindset interventions that aim to lessen gender differences in competence-related beliefs (e.g., Blackwell et al., 2007; Lee et al., 2021; Yeager et al., 2019). Lee and colleagues (2021) summarized different strategies used in interventions to promote math competence-related beliefs, including communicating that individuals' skills and intelligence are not fixed and can develop, and that the brain is malleable. However, these intervention studies are more recent and have been conducted over short periods of time—making it unclear if the intervention effects are susceptible to fade out over time (Bailey et al., 2020). Moreover, interventions are not widely implemented, and NASEM (2020) appropriately highlights that the underrepresentation of women in STEM is a systemic problem that needs to be addressed through policy changes. Though interventions aimed at reducing gender disparities in students' math competence-related beliefs have been shown to be effective (see Lee et al., 2021; Levene & Pantoja, 2021), these efforts are isolated and do represent widespread policy changes. Thus, we expected gender differences in students' math competence beliefs to persist across historical time.

Gender Differences Within Racial/Ethnic Groups - An Intersectional Lens

Gender and race/ethnicity have often been examined separately in research, but they are not isolated factors and codetermine individuals' beliefs (Else-Quest & Hyde, 2016; Hyde, 2007; Parker et al., 2020). An intersectional approach captures the idea that individuals' experiences within one social identity are influenced by their other social identities (Crenshaw, 2019). Thus, examining how the intersection of gender and race/ethnicity explains students' competence-related beliefs is crucial. For example, even though male students, on average, are expected to hold more privilege than female students in STEM, this pattern may not be universal. We can ask what happens if an individual simultaneously experiences a positively stereotyped social identity and negatively stereotyped social identity (i.e., an

African-American male student whose race/ethnicity but not gender is marginalized in STEM, see Hsieh, Simpkins & Eccles, 2021).

So why might gender differences in students' competence-related beliefs vary by race/ethnicity? Racial/ethnic stereotypes and discrimination impact motivation and could weaken or exacerbate marginalization due to one's gender. For example, the model minority stereotype for Asian-Americans includes the assumption that they excel in math, which could lessen the difference between male and female Asian-American students as both are expected to exhibit high math achievement (Trytten et al., 2012). Moreover, Latinx-Americans and African-Americans are more likely to face discrimination based on English fluency, social class, and skin color in the U.S. (Alfaro et al., 2009; Rosenbloom & Way, 2004). Two U.S. studies found that African-American girls felt adults underestimated their STEM abilities both due to their gender and race (Archer et al., 2015; Bruning et al., 2015). These racist and sexist experiences for these groups invariably shape their perceptions about their academic abilities and success in STEM (Eccles & Wigfield, 2020); both male and female students of these racial/ethnic groups may get marginalized in STEM thereby lessening gender differences in each of these groups.

Though gender differences have been examined extensively, very little is known about whether gender differences replicate in each of the four largest racial/ethnic groups in the U.S. Some initial findings suggest gender differences do not necessarily replicate across all races/ethnicities (Hsieh et al., 2021; Parker et al., 2020). For example, gender differences favoring male high school students in math ability self-concepts have emerged for European-Americans and Latinx-Americans (Else-Quest et al., 2013; Seo et al., 2019); however, the findings were less consistent for African-Americans and Asian-Americans. Though Seo and colleagues (2019) reported no gender differences among African-American and Asian-American high school students in 11th grade, Else-Quest and colleagues (2013) found gender differences in math ability self-concepts favoring male African-Americans and Asian-Americans throughout high school. The gender differences were larger for Asian-Americans (d = .41) than African-Americans (d = .16) (Else-Quest et al., 2013). More research is needed to address these limited, inconsistent findings.

The importance of replication in psychology

One issue confronting the discipline is that empirical results have often not been replicated (Duncan et al., 2014; Plucker & Makel, 2021). Replication studies, particularly those that use multiple datasets to check the accuracy and robustness of the effect sizes, are critical but rare (Duncan et al., 2014). Conceptual replication, where researchers use data that includes similar concepts based on divergent methods (e.g., different measures, contexts, participant characteristics), is beneficial for testing whether a theoretical perspective holds up across different cultures or historical time points (Plucker & Makel, 2021).

Replicating empirical results is imperative to account for potential differences across multiple datasets, e.g., the historical background or participant characteristics.

Theorists argue that motivational processes are shaped by these indicators (Eccles & Wigfield, 2020); however, most studies have explored psychological processes only among European-American, middle-class samples. We see the advantage of using datasets from different historical times. Advantages lay, for example, in examining possible changes in psychological processes over time (historical changes), using existing datasets for secondary analyses and not always collecting new datasets and over-researched samples, and also the theoretically assumed situatedness in psychological processes in different contexts (see Eccles & Wigfield, 2020). Even current datasets have the advantage of providing contemporary results, it is also important to test these processes across historical time. In our study, we aim to use datasets from the 1980s to 2010s to test the replication of math gender differences overall and within the four largest U.S. racial/ethnic groups.

The present study

This study is a conceptual replication investigating gender differences in high school students' math competence-related beliefs across six U.S. datasets collected over 30 years. In addition to replicating these gender differences across the six datasets, we designed this study to be a conceptual replication of the Else-Quest et al.'s (2013) study of gender differences (within racial/ethnic groups) in math competence-related beliefs. Though we purposefully used the same analytic methods, namely analysis of variance, our study used prior achievement as a covariate, focused only on students' math competence-related beliefs as the dependent variable, and utilized multiple datasets (see a full comparison of our study to Else-Quest et al. (2013) in the Supplemental Material, Part F).

The goals of our study were to test the extent to which gender differences within each racial/ethnic group replicated across the six datasets included in the current study and replicated the difference those found by Else-Quest et al. (2013). Additionally, our goal also was to extend prior research by testing potential historical changes in those differences. The systematic approach of our study will help uncover gender differences and potential variation in those gender differences across and within racial/ethnic groups. Furthermore, this study will provide insight into whether such differences vary as a function of when the data were collected.

The following research questions were examined:

(RQ1) What are the overall gender differences in high school students' competence-related beliefs in math from 9th to 12th grade? Does the size of gender differences differ systematically across datasets and indicate historical changes?

(RQ2) What are the gender differences in high school students' competence-related beliefs in math from 9th to 12th grade within each racial/ethnic group? Does the size of the gender differences within each racial/ethnic group differ systematically across datasets and indicate historical changes?

Overall, we expected male students to report higher math competence-related beliefs than female students in 9th to 12th grade (Else-Quest et al., 2013; Parker et al., 2020). Despite cultural changes in gender roles, we expected no historical changes in gender differences (Parker et al., 2018; Scherrer & Preckel, 2019).

Based on prior research, we expected these gender differences to emerge within European- and Latinx-Americans (Else-Quest et al., 2013; Seo et al., 2019). Due to the existing inconsistent findings on gender differences within African- and Asian-Americans (Else-Quest et al., 2013; Seo et al., 2019), we do not have a priori expectations of the gender differences within these groups. We think it is important to acknowledge that by examining students' race/ethnicity as a potential factor of interindividual differences in motivational beliefs from a social-historical perspective, we do not assume biological differences between races/ethnicities in motivational beliefs (Urdan & Bruchmann, 2018).

We used multiple background variables as covariates, including prior achievement, parent education, and financial background. In order to understand when gender differences may begin to affect math competence-related beliefs in high school, it is crucial to compare effects controlling for students' background. Thus, we investigated gender differences (within racial/ethnic groups) while holding achievement and family social-economic background constant throughout the analyses (see Chancer & Watkins, 2006).

METHOD

Utilized data

In this study, we utilized six longitudinal datasets: the California Achievement Motivation Project (CAMP), the Childhood and Beyond Study (CAB), the Maryland Adolescents Development in Context Study (MADICS), the Panel Study of Income Dynamics Study-Child Development Supplement (PSID-CDS), the Michigan Study of Adolescent and Adult Life Transitions (MSALT), and the High School Longitudinal Study (HSLS). All of these datasets include math competence-related beliefs constructs during high school measured with the same or similar items affording tests of conceptual replication (Plucker & Makel, 2021). These datasets vary in their design, such as the grade levels included, data collection years, the number of cohorts, and participant demographics (see for orientation Figure 1). We utilized all high school data available in each of these datasets. We capitalized on the rich variability across these datasets to examine the extent to which gender differences in students' math competence-related beliefs replicated. An overview of these datasets is below and provided in Supplemental Materials, Part A, along with descriptive statistics.

Participants

For all datasets, students were included if they (a) were in high school (9th to 12th grade) and had complete data on (b) their race/ethnicity, gender, grade level, family socioeconomic background, parental education, and achievement, and (c) math competence-related beliefs. Across the datasets, the total sample consisted of 24,280[1] students (49.5% male students) with 11% African-, 9% Asian-; 30% Latinx- and 50% European-Americans. A subsample of 23,070[1] students (49.5% male students) with 11% African-, 9% Asian-; 32% Latinx- and 48% European-Americans were used to answer research question 2; data from CAB and MSALT were not include in research question 2 because more than 95% of students in these datasets were European-Americans. Descriptive statistics for each dataset are in Supplemental Material, Part A; information on missing data is in Supplemental Material, Part D. Below, we describe the participants in each dataset.

MSALT

MSALT is a longitudinal dataset with mostly European-American students from working and middle-class communities in Michigan. Students were surveyed from 1983 to 2000. For this study, a subsample of 789 students from Wave 5 (1988/89, 10th grade) and Wave 6 (1990, 12th grade) were included (49% male students). We included 402 students from 10th grade and 387 students from 12th grade. Only European-American students were included in this study as students from other races/ethnicities were underrepresented (5%). The annual income ranged from $29,999 or less (29%) to over $40,000 (42%). Approximately 70% of the parents had at least some college or technical degree.

CAB

CAB primarily included European-American students and their parents from lower-middle to middle-income families in Southeastern Michigan. It is a longitudinal, cohort-sequential study surveying each cohort over several years in school from 1986 to 1999. For this study, a subsample of 430 students with data collected from 1994 to 1999 were included (47% male students). There were 111 students in 9th grade, 155 in 10th grade, 55 in 11th grade, and 109 in 12th grade. Only European-American students were included in this study as students from other races/ethnicities were underrepresented (5%). Participating families had an annual income ranging from $39,999 or less (7%) to over $80,000 (28%). Approximately 92% of the families had one parent with at least a college degree.

MADICS

MADICS is a longitudinal dataset with mostly African-American (64%) and European-American (36%) students from Maryland. Students represented diverse socioeconomic backgrounds and were surveyed between 1991 and 2000. For this study, a subsample of 690 students from Wave 4 (collected in 1996/1997, 11th grade) were included (50% male students). Participating families had an annual income ranging from $39,999 or less (25%) to over $80,000 (26%). Average parental education was 15 years (*SD* = 2.71), which is equivalent to some college or a college degree.

---

[1] All HSLS-values and values including HSLS-values (i.e. total sample sizes) were rounded to the nearest tens place per IES restricted-use guidelines.

PSID-CDS
PSID-CDS is a multi-cohort nationally representative longitudinal dataset of families and their children from 1997 to 2019. For this study, a subsample of 1,264 students were included: cohorts 6-8 in 2002 and cohorts 11-13 in 2007 (50% male students). There were 354 students in 9th grade, 327 in 10th grade, 323 in 11th grade and 260 in 12th grade. Approximately 46% of the sample was African-American and 54% European-American. The annual family income ranged from $35,000 or less (31%) to over $95,000 (29%). The average parental education was 13.8 years (*SD* = 2.12), which is equivalent to some college or a college degree.

CAMP
CAMP is a longitudinal, cohort-sequential dataset of middle and high school students from four school districts in Southern California. Students were surveyed in their math classes over 2 years, between 2004 and 2006. A subsample of 6,540 students were included (49% male students; 14% Asian-Americans, 12% European-Americans, 74% Latinx-Americans). There were 2,721 students in 9th grade, 1,742 in 10th grade, 1,563 in 11th grade, and 514 in 12th grade. Among them, 59% of students were eligible for free/reduced lunch (i.e., an indicator for family SES), and 20% of the parents had at least some college or a technical degree.

HSLS
HSLS is a nationally representative longitudinal dataset of high school students in 9th (2009) and 11th grade (2011). For this study, a subsample of 14,570[1] students from both waves were included (50% male students). We included 6,690[1] students from 9th grade and 7,880[1] students from 11th grade. The sample included approximately 10% African-Americans, 8% Asian-Americans, 65% European-Americans, and 17% Latinx-Americans. The average annual income ranged from $35,000 or less (26%) to over $95,000 (31%), and 60% of the parents had at least some college degree.

Instruments

Competence-related beliefs in math
Students' competence-related beliefs were operationalized using students' ability self-concepts in most of the surveys (CAB, MSALT, PSID-CDS, HSLS, MADICS) and self-efficacy in one survey (CAMP). They were assessed once per academic year except for the CAMP dataset, in which they were measured twice within each academic year. To match the measurement points, we averaged the two measures in CAMP during the academic year to get an average score for each grade level.

The authors carried out psychometric analyses in each dataset to examine the properties of all scales. The competence-related belief scales included two to five items, with factor loadings of $.53 \leq \lambda \leq .96$ and internal consistency of $.84 \leq \omega \leq .94$. For MADICS, the Spearman-Brown reliability coefficient was calculated because the scale included only two items (Eisinga et al., 2013). For a full list of items, internal consistency, factor loadings and response scale by dataset, see Supplementary Material, Part B.

In each dataset, measurement invariance analyses across gender, race/ethnicity and for longitudinal studies across time*gender and time*race/ethnicity were conducted to examine if adolescents' math competence-related beliefs had similar measurement properties (Corral & Landrine, 2010). The models evidenced full configural invariance, full metric invariance, and full or partial scalar invariance for all measures in each dataset (see Supplemental Material, Part C).

Background variables
To obtain information on students' gender and race/ethnicity, we used self-reports (CAB, MADICS, MSALT, HSLS, and PSID-CDS) and district data (CAMP). Below, we describe the covariates. Frequencies of the background variables for each dataset are described in the Supplemental Material, Part A.

Achievement. Students' math achievement is associated with their competence-related beliefs (Pietsch et al., 2003). Thus, we included an indicator of prior math or general academic achievement. The indicators were (a) district-reported math exam scores (the California Standards Test in CAMP and the Michigan Education Assessment program math test scores in MSALT); (b) standardized test scores (the Woodcock-Johnson Revised Test of Achievement in math in PSID-CDS and algebraic math assessment in HSLS), (c) general grade point average in MADICS and (d) students' general intelligent quotient in CAB.

Parent education. Previous studies indicate that parents' educational background impacts their children's STEM competence-related beliefs (Simpkins, Fredricks & Eccles, 2015). In the six datasets, parents' education was operationalized using (a) the parents' highest education degree (CAB, CAMP, HSLS, and MSALT) or (b) the highest number of years of formal schooling of parents (MADICS and PSID-CDS).

Family financial background. Families' financial background is a factor that influences students' math competence-related beliefs (Parker et al., 2020). The family financial background was operationalized using information on (a) family income (CAB, HSLS, MSALT, PSID-CDS, and MADICS) or (b) student participation in the school lunch program (CAMP).

Statistical analysis

This study aimed to examine gender differences in students' math competence-related beliefs during high school. In line with previous research, a two-step approach was used to fully investigate gender differences by testing (a) overall gender differences (step 1), and (b) gender differences within each of the four racial/ethnic groups (step 2, see Else-Quest et al., 2013).

It is crucial to test gender differences with and without covariates. This dual approach allowed us to estimate the size of the differences between male and female students' math competence-related beliefs (*What are the gender differences in each dataset?*) as well as the size of the differences after accounting for important background variables (*What are the gender differences with students'* *demographic background held constant in each dataset?*). For simplicity, we describe the results with covariates in this paper. Results without covariates are presented in the Supplemental Material, Part E. We briefly compare the results concerning gender differences without and with covariates in the last section of the Results section.

The datasets varied in methodological characteristics, including sampling methods, sample size, response scales, number of items, and study design. These differences necessitated conducting all analyses separately for each dataset (Hedges & Vevea, 1998). For instance, all analyses for PSID-CDS and HSLS were adjusted for the complex sampling design (i.e., sampling weights [PSID-CDS, HSLS], primary sampling unit [HSLS], and strata [HSLS]). After all of the analyses were estimated for each dataset, we compared effect sizes weighted by sample size across datasets. We explain the analyses within and across datasets in more detail below.

Analyses within each dataset

We used SPSS version 26 to estimate the ANCOVAs. For HSLS and PSID-CDS, we used the complex samples module. As explained earlier, we tested gender differences through two main steps (step 1: overall gender differences, step 2: gender differences with each racial/ethnic group) with two sub-analyses within each step (analyses without and with covariates). In total, we conducted 17 ANCOVAs estimating overall gender differences and 30 ANCOVAs estimating gender differences with each racial/ethnic group with covariates; we also estimated a parallel set of 47 ANOVAS that did not include covariates. Analysis of variance techniques were utilized so the means of all sub-groups could be compared with each other (Del Río-González et al., 2021). Each dataset included different grade levels (see Figure 1). Thus, the analyses were conducted separately at each grade level in each dataset. We analyzed gender differences with cross-sectional data in each dataset to compare results across datasets. Because most of the datasets included longitudinal data in high school (except MADICS), we used SPSS to randomly assign students to one grade level. Thus, data from each student was included only once at a randomly assigned grade level, meeting the analysis of variance assumption for independence. The participant descriptions presented earlier in the Method section already took this random selection into account.

Effect sizes (Hedges' g) at each grade level in each dataset were used to calculate the combined (weighted average) effect size of the gender differences. Hedges' g corrects for biases associated with small sample sizes (Hedges & Olkin, 1985). For effect sizes, it is mandatory to interpret the results concerning the topic of interest, the data and project design, and the study aims (see Bakker et al., 2019). Guided by previous research (Else-Quest et al., 2010; Parker et al., 2018), we refer to meaningful gender effects with effect sizes $g \geq .15$. However, we do not use these as strict cut-offs (e.g., effect sizes of .14 and .15 both were interpreted as meaningful). Furthermore, because different Likert scales were used across datasets (1-7 [MSALT, CAB, MADICS, PSID-CDS]; 1-5 [CAMP]; 1-4 [HSLS]), we used the transformation of mean scores via Percent of Maximum Possible (POMP) score to descriptively compare means across datasets (Cohen et al., 1999). The POMP score transforms different Likert scales on one common scale from 0 (minimum) to 100 (maximum) using the following formula: POMP = [(observed score - minimum possible score)/(maximum possible score - minimum possible score)] × 100.

Analyses across datasets
After estimating gender differences (overall and within each racial/ethnic group) in each dataset, we tested the heterogeneity of the effects across datasets using Comprehensive Meta-Analysis version 3.3 (CMA, Borenstein et al., 2010). We estimated fixed and random effects and compared both models using the conventional heterogeneity tests in meta-analyses based on Q statistics and the I^2 (see Hedges & Schauer, 2019). The fixed effect model assumes the homogeneity of effects, while the random effects model assumes the heterogeneity of effects. Thus, Q-statistics test the null hypothesis of no meaningful differences in the effect sizes across the included studies (Hedges & Schauer, 2019).

RESULTS

This section reports results on gender differences overall (RQ1) and gender differences within each racial/ethnic group (RQ2). In each section, we describe the effect sizes at each grade level across the datasets. Afterward, we report results on the heterogeneity of the gender differences across datasets. Several datasets from different historical times are used and, thus, allowed us to test the heterogeneity of gender differences to address historical differences.

For simplicity, we focus on results with covariates. Results comparing gender differences without and with covariates are briefly mentioned at the end of this section (see supplementary material for fully reported results without covariates).

Gender differences overall (RQ1)

Descriptive information on students' math competence-related beliefs by grade level and by gender is presented in Table 1. All reported effect sizes in this section refer to the combined effects calculated across datasets. Results of the ANCOVAs with effect sizes computed on each dataset as well as indicators of the heterogeneity of these effects across datasets are represented in Tables 2 and 3. Patterns of gender differences for students' math competence-related beliefs are plotted in Figure 2.

At all grade levels, meaningful gender differences emerged that favored boys even after controlling for covariates ($.14 \leq g \leq .23$; Tables 2 and 3). However, the combined gender effects at 9th and 11th grade were weak with $g = .14$. The combined effects at 10th and 12th grade were small with $g = .22$ and $g = .23$. The heterogeneity tests indicated that there was some variation in the effect sizes across datasets at 9th grade but not at 10th through 12th grade (see the Q-statistics in Tables 2 & 3). As shown in Figure 2, the variability in the effect sizes at 9th grade was not related to the survey year of the dataset. Thus, there was no evidence that gender differences varied based on the year data were collected.

Gender differences within each racial/ethnic group (RQ2)

Descriptive information on students' math competence-related beliefs by grade level, and by gender and race/ethnicity is presented in Table 4. Reported effect sizes in this section refer to either the combined effects calculated across datasets or effect sizes that result only from one dataset; combined effects could not be calculated at certain grade levels for Latinx-, Asian-, and African-Americans because only one dataset included that racial/ethnic group at some grade levels (e.g., data for Latinx 10th graders were only available in CAMP; see Figure 1). Results of ANCOVAs with effect sizes computed on each dataset as well as indicators of the heterogeneity of these effects across datasets are represented in Tables 5 and 6. In alphabetical order, we report gender differences for African-, Asian-, European- and Latinx-Americans.

Gender differences for African-Americans

The gender differences for students' math competence-related beliefs controlling for covariates within African-Americans are plotted in Figure 3. We report the combined effect size (g) for African-Americans at 9th and 11th grade and the effect size (g) from one dataset at 10th and 12th grade (Tables 5 and 6). At 9th to 11th grade, no meaningful gender differences emerged (9th: g = .00; 10th: g =-.12; 11th: g = .10). At 12th grade, the effect sizes indicated meaningful gender differences (g = -.15) that favored girls. The test for heterogeneity could be calculated at 9th and 11th grade; the findings suggest that the size of the gender differences was similar across datasets at each grade level (see the Q-statistics in Tables 5 and 6). Thus, there was no evidence that gender differences varied across datasets for African-Americans at 9th or 11th grade.

Gender differences for Asian-Americans

The gender differences for students' math competence-related beliefs within Asian-Americans with covariates are plotted in Figure 4. We report the combined effect size (g) for Asian-Americans at 9th and 11th grade and the effect size (g) from one dataset at 10th and 12th grade. At every grade level, meaningful gender differences emerged that always favored boys (9th: g = .39; 10th: g = .38; 11th: g = .19; 12th: g = .62). The test for heterogeneity could be computed at 9th and 11th grade and indicated that the size of the gender differences was similar across datasets at each grade level (see the Q-statistics in Tables 5 and 6). Thus, there was no evidence that gender differences varied across datasets for Asian-Americans at 9th or 11th grade.

Gender differences for European-Americans

The gender differences for students' math competence-related beliefs within European-Americans with covariates are plotted in Figure 5. We report the combined effect size (g) for European-Americans at all grade levels. At every grade level, meaningful gender differences emerged that always favored boys ($.17 \leq g \leq .44$). The test for heterogeneity was computed at each grade level and indicated that the size of the gender differences among European-Americans was similar across datasets from 9th to 12th grade (see the Q-statistics in Tables 5 and 6). Thus, there was no evidence that gender effects varied across datasets for European-Americans.

Gender differences for Latinx-Americans

The gender differences for students' math competence-related beliefs within Latinx-Americans with covariates are plotted in Figure 6. We report the combined effect size (g) for Latinx-Americans at 9th and 11th grade and the effect size (g) from one dataset at 10th and 12th grade. At every grade level, meaningful gender differences emerged that always favored boys (9th: g = .20; 10th: g = .28; 11th: g = .17; 12th: g = .18). The test for heterogeneity could be tested at 9th and 11th grade; the findings indicated that the size of the gender differences was similar across datasets at 9th and 11th grade (see the Q-statistics in Tables 5 and 6). Thus, there was no evidence that gender effects varied across datasets for Latinx-Americans.

Results on gender differences without covariates

The analyses testing gender differences without covariates (overall and within each racial/ethnic group) are reported in the Supplemental Material, Part E. We would like to point out two findings, which became evident in the comparison of gender differences within racial/ethnic groups without and with covariates.

First, for African-Americans, the combined effects were slightly stronger when parceling out the control variables than when the control variables were not included (-.06 ≤ g ≤ .18). Specifically, the gender differences were not significant when we did not include the covariates. Second, the heterogeneity of effects across datasets was larger when we did not include the covariates, i.e., across races/ethnicities (RQ1) and Asian-Americans (RQ2). This pattern suggests that the comparability of effect sizes between datasets is better achieved after controlling for the covariates.

DISCUSSION

Gender, race, and ethnicity have been found to explain variation in students' motivational beliefs across K-12 grades (Zusho & Kumar, 2018). The resulting interindividual differences are particularly profound in STEM (Else-Quest et al., 2013; Parker et al., 2018; Parker et al., 2020). This study investigated the gender differences in math competence-related beliefs among African-, Asian-, European-, and Latinx-American students from 9th to 12th grade, and also across six datasets between the 1980s and 2010s to test if these findings replicate across historical time.

Replication of gender differences overall

Supporting our hypothesis, we found that male students had higher math competence-related beliefs than female students in high school. This pattern was consistent from 9th to 12th grade across datasets though the gender differences were weak at 9th and 11th grade with effect sizes of g = .14. The direction and size of these gender differences align with prior work by Else-Quest and colleagues

(2013) and other studies that sampled high school students from the 1980s to 2010s (Graham & Morales-Chicas, 2015; Nagy et al., 2010; Parker et al., 2020; Watt et al., 2012). These results need to be interpreted in the context of gender equality in students' math achievement (see McGraw et al., 2006). Explanations for persistent gender differences in students' competence-related beliefs despite equality in math achievement are that boys and girls are exposed to differing socialization experiences (e.g., interactions with parents, friends, and teachers) that can inform their perceptions of their competences. Research on gender roles indicates that gender-role stereotypes and beliefs (e.g., male students are good at math, value money, and working with tools; female students value helping others) influence parents' opinions of their children's competence to perform various activities (Eccles et al., 1990). Simpkins and colleagues (2015) found that parents' beliefs about their children's competence and the extent to which parents provide support is influenced by their children's gender, and is expected to shape children's motivational beliefs.

Replication of gender differences within racial/ethnic groups

Prior research findings on gender differences within different racial/ethnic populations are limited and mixed, with some reporting gender differences and others finding no differences (Else-Quest et al., 2013; Seo et al., 2019). Supporting our hypothesis and replicating findings from Else-Quest et al. (2013), we identified gender differences in math competence-related beliefs favoring male students among European- and Latinx-Americans that replicated across datasets and all high school grades. These observed gender differences within European- and Latinx-Americans are consistent with previous research (Else-Quest & Hyde, 2016; Seo et al., 2019) and the overall trend of gender differences in this study favoring male students. We also found gender differences favoring male students among Asian-Americans that support previous work from Else-Quest and colleagues (2013) and the overall trend found in this study. Thus, we replicated that gender differences emerged only for certain racial/ethnic groups in high school (i.e., European-, Latinx-, and Asian-American) and always favored boys in these groups.

Our findings for Asian-Americans add to theoretical assumptions on math stereotypes. We know about stereotypes of high expectations of math achievement for both male and female Asian-Americans. A combination of scholarship on stereotype threat in math (e.g., Steele, 2013) and model minority stereotypes (Trytten et al., 2012) suggest that gender gaps might be smaller for Asian-Americans. However, our results, along with prior empirical studies by Else-Quest et al. (2013), point to gender differences among Asian Americans that favor boys with large effect sizes. Cumulatively, it is possible that Asian-American students have higher competence beliefs in math compared to other racial/ethnic groups as suggested by multiple theories, but this group also exhibits large gender differences as suggested by empirical evidence. It is also possible that gender differences might be more pronounced among students who exhibit high math competence or high math competence beliefs, though this needs to be systematically tested (see Baye & Monseur, 2016). Studies are needed on how psychological processes and societal beliefs, such as stereotypes, are related to gender differences within racial/ethnic groups.

The findings for African-Americans differed from the overall trends and the three other racial/ethnic groups. Specifically, male and female African-American students did not differ in their math competence-related beliefs from 9th to 11th grade - a finding that replicated across studies. Though these non-significant differences align with findings from Seo and colleagues (2019), it varies from results by Else-Quest and colleagues (2013) and differs from our overall trend of gender differences favoring male students. One possible explanation is that, compared to European-American youth and their families, African-American youth and their families are less likely to endorse stereotypes privileging boys in math and/or may hold higher academic expectations for their daughters (Evans et al., 2011; Rowley et al., 2007; Skinner et al., 2021; Wood et al., 2010). By communicating high academic expectations for their daughters and disavowing traditional math gender stereotypes, African-American parents may set their daughters up to have similar math competence-related beliefs as their male peers. Another reason for different results than Else-Quest et al. (2013) might be that we controlled for prior achievement. Thus, we exclude variance of math competence-related beliefs that is explained by prior achievement. It might be that we did not find gender differences for African-Americans controlling for achievement, because achievement might be a more crucial predictor of students' competence-related beliefs for African-Americans in high school than gender.

In addition, we also found gender differences favoring female students over male students in 12th grade for African-Americans; this result contradicts our study's overall gender differences favoring male students in 12th grade. The finding on gender differences favoring African-American female students in 12th grade should be interpreted cautiously as these observed differences were limited to only one dataset (PSID-CDS). But, the findings complement other research indicating that female minority students persist further in math (Safavian, 2019) and male African-American students experience more discrimination than their peers (Feliciano, 2012). However, it is important to note that the findings for African-American students often were different than the overall trends in this study and work by Else-Quest et al. (2013). Thus, more research among African-American populations is warranted to better understand how findings on gender equality in math competence-related beliefs might reflect the empowerment of female students in STEM or structural and systemic barriers in STEM for male students. In general, more research of this kind is needed to examine trends across grades. It might be important to investigate why gender differences occurred for Asian-, European- and Latinx-Americans but not African-Americans in 9th to 11th grade.

Historical changes in gender differences

Supporting our hypotheses, we found no evidence of historical changes concerning gender differences in students' math competence-related beliefs using six U.S. datasets from the 1980s to the 2010s. Our results align with prior findings in meta-analyses indicating no effect of the publication year or data collection year on students’ motivational beliefs, motivational development, and relevant gender effects (Parker et al., 2018; Scherrer & Preckel, 2019).

Our findings make two significant contributions. First, researchers have debated the situatedness of students’ motivational development (Eccles & Wigfield, 2020), as systemic changes in society and policy might influence motivational beliefs and gender disparities. Over the last few decades, several initiatives have been launched to increase gender equality in terms of college degrees in STEM majors, raise the recruitment and retention of women in STEM, and counteract the achievement gap (see NASEM, 2020; NSF, 2021). There also have been multiple meaningful interventions that support marginalized students (e.g., teacher-level interventions, interventions on creating inclusive environments, student-level interventions), some of which have had positive impacts on students’ motivational beliefs in STEM (for reviews of these interventions, see Lee et al., 2021; Levene & Pantoja, 2021; Rosenzweig & Wigfield, 2016).

We asked whether the patterns of gender differences in students’ motivational beliefs might differ across historical contexts. Our study demonstrates that gender differences in math competence-related beliefs of high school students were stable across datasets from the 1980s to the 2010s. Our finding is quite important given the amount of effort that has been devoted to reducing or eliminating gender differences in STEM over the last several decades. However, there is an important nuance we need to note. We did not test the effectiveness of interventions aimed to counteract gender differences over time, but used datasets of different historical times to examine gender differences in math competence-related beliefs.

Stable gender differences across the last several decades could be explained by the fact that some of the interventions aiming to reduce gender disparities are comparatively recent, smaller in scale and thus are not reflected in the data used in the study. Moreover, any type of intervention, whether student-centered, fostering students’ attitudes towards STEM, or more general policy-related initiatives aimed to create inclusive environments for women in STEM, will need to be implemented widely to create fundamental change in the broader population. Particularly student-centered interventions aimed at improving students’ perception of their own capabilities in STEM showing short-term success so far, have not been implemented widely until very recently (Bailey et al., 2020). Ultimately, some of the more systemic changes happening in the most recent decades are likely to take time to take effect and would not have shown up in the data used in the current study. More importantly, the development of competence-related beliefs is a highly complex process that is influenced by many different factors, including but not limited to experiences in educational settings as well as their interactions with family members, friends and the expectations and norms they communicate. In addition, societal norms are communicated through media that adolescents engage with on a daily basis. Just changing one of these influences might not be sufficient to create long-term change.

Ultimately, STEM education is biased and has led to decades of unequal opportunities for those who are marginalized, such as women and Latinx-Americans

(Honey et al., 2020). Our study provides important information regarding the historical stability of gender differences in adolescents overall. However, we were more limited in our ability to investigate historical changes of gender differences within each racial/ethnic group (e.g., the three datasets with information on African-Americans span 15 years from 1996 to 2011). We encourage future research to examine historical changes in gender differences among multiple racial/ethnic groups using datasets from multiple historical decades.

Limitations and future directions

Though these findings make several contributions, the limitations must be considered. First, the current study investigated mean-level differences in math competence-related beliefs across datasets and racial/ethnic groups. We, however, do not want to create a misperception that these racial/ethnic groups are monolithic (Urdan & Bruchmann, 2018). Assessing race/ethnicity and gender using categorical variables simplifies the complexity of culture and ethnicity/race and negates the existing the rich variability within each group (Zusho & Kumar, 2018). Future research should explore the potential variation that exists within each of the investigated subgroups (for example, Hsieh et al., 2021).

Second, given that the current study used existing datasets, the measures were not identical across all studies (CAMP: self-efficacy; all other datasets: ability self-concept). Similar to the approach used in meta-analyses (see Parker et al., 2020), however, the focus of the current study was the comparison at the construct level, which is a test of conceptual replication. Research has shown that self-efficacy and self-concept measures are highly similar and not indistinguishable empirically at times (Wigfield & Eccles, 2002).

Third, we excluded all cases with missing data (listwise deletion). This was done because we used analysis of variance and a meta-analytic approach to compare mean-level differences across multiple datasets. One challenge with analysis of variance is handling missing data. Though multiple imputation is one way of dealing with missing data, it does not work well with the analysis of variance techniques (see Graham, 2012; Finch, 2016). However, scholars have shown that the results obtained with analysis of variance and listwise deletion are close to the results found with other ways of handling missing data (see Grund et al., 2016). We, therefore, excluded cases with missing data though it is possible that this decision might lead to biased results as excluded cases were different than included cases in some datasets (see Supplemental Material, Part D). We would like to highlight that we replicated our findings across 6 U.S. datasets that are in line with other results based on regression analyses that estimated missing data (see Arens et al., 2017, Marsh et al., 2021), which would suggest that our findings are less likely to be fundamentally biased.

Also, our paper conceptualizes gender as either identifying as female or male because the utilized datasets measured gender as a female/male binary. However, gender is more complex. Future research might include other gender identities (such as non-binary and other transgender identities) as well as include more detailed measures to capture the multiple facets of gender identity, such as centrality and felt typicality (Tobin et al., 2010).

We envision two further directions for this line of research. First, given the interplay of expectancy and value beliefs, it is worthwhile to study historical trends in gender differences in math subjective task values. Throughout history, society's projected needs have had implications for the agendas of our social institutions, economic growth, and technological advancements. Gender role attitudes are shifting in tandem with changes in historical context. This also impacts messages about what vocations are important, and hence directs what subjects of study will be relevant in the future in terms of labor force and social importance. These broad values shape what gender groups consider to be worthwhile investments of their resources. This also inevitably impacts the implicit and explicit messaging that children receive and internalize regarding the importance and utility of math, as well as molding their experiences and engagement with math through the provision of activities, experiences, and resources.

Second, our study included parent education and financial background as covariates. However, we know about the intersection of gender, race/ethnicity, and socioeconomic background (class) (see Chancer & Watkins, 2006). As seen in our paper, gender differences in math competence-related beliefs were found for European-, Asian-, and Latinx-American students, but not for African-American students. Such patterns have also been found for performance in STEM subjects (McGraw et al., 2006). Building on these findings, other research found that these gender differences were more pronounced for students from lower socioeconomic backgrounds (Entwisle et al., 2007). This relationship could also apply to competence-related beliefs. One challenge when testing replication of the intersection of these factors is the difference of the operationalization of socioeconomic background and class across datasets. However, future research can consider a way to account for socioeconomic background as a predictor that is interrelated with gender and race/ethnicity.

Conclusion

In sum, our findings have several implications. This study adds empirical value given the relevance of using an intersectional lens to more accurately capture the interindividual differences in our increasingly diverse society (Urdan & Bruchmann, 2018). For instance, we found gender differences in high school students’ math competence-related beliefs among Asian-, European- and Latinx-Americans but not in African-Americans. Therefore, one empirical implication is that future studies need to take both gender and race/ethnicity into account when aiming to understand students’ STEM motivational beliefs. Furthermore, aligned with the recommendations by Plucker and Makel (2021), we highlight the importance of replication in motivational psychology. We demonstrate the need for scholars to (a) use multiple datasets to replicate findings and identify potential differences in the patterns, and (b) use datasets from different decades to understand the historical changes in motivational processes.

References

Alfaro, E. C., Umaña-Taylor, A. J., Gonzales-Backen, M. A., Bámaca, M. Y., & Zeiders, K. H. (2009). Latino adolescents' academic success: The role of discrimination, academic motivation, and gender. *Journal of Adolescence*, *32*(4), 941–962. https://doi.org/10.1016/j.adolescence.2008.08.007

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & Vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. Journal of Educational Psychology, 109(5), 621.

Archer, L., DeWitt, J., & Osborne, J. (2015). Is science for us? Black students' and parents' views of science and science careers. Science Education, 99(2), 199-237. https://doi.org/10.1002/sce.21146

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. Journal of Research in Science Teaching, 47(5), 564-582. https://doi.org/10.1002/tea.20353

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. Psychological Science in the Public Interest, 21(2), 55-97.

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. Large-scale Assessments in Education, 4(1), 1-16.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. Educational psychology review, 15(1), 1-40.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender Diversity in STEM Disciplines: A Multiple Factor Problem. *Entropy*, *21*(1), 30. https://doi.org/10.3390/e21010030

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(49), 31063–31069. https://doi.org/10.1073/pnas.2008704117

Bruning, M.J., Bystydzienski, J., & Eisenhart, M. (2015). Intersectionality as a framework for understanding diverse young women's commitment to engineering. Journal of Women and Minorities in Science and Engineering, 21(1), 1-26. https://doi.org/10.1615/JWomenMinorScienEng.2014007345

Cheryan, S., Lombard, E. J., Hudson, L., Louis, K., Plaut, V. C., & Murphy, M. C. (2020). Double isolation: Identity expression threat predicts greater gender disparities in computer science. Self and Identity, 19(4), 412-434. http://dx.doi.org/10.1080/15298868.2019.1609576

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The Problem of Units and the Circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Crenshaw, K. (2019). 'Difference' through intersectionality. In S. Arya & A. S. Rathore (Eds.), *Dalit feminist theory: A reader* (pp. 139–149). Routledge.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. Journal of research in personality, 19(2), 109-134.

Del Río-González, A. M., Holt, S. L., & Bowleg, L. (2021). Powering and Structuring Intersectionality: Beyond Main and Interactive Associations. *Research on Child and Adolescent Psychopathology*, *49*(1), 33–37. https://doi.org/10.1007/s10802-020-00720-w

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*(11), 2417–2425. https://doi.org/10.1037/a0037996

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Else-Quest, N. M., & Hyde, J. S. (2016). The Authors Respond. *Psychology of Women Quarterly*, *40*(3), 351–352. https://doi.org/10.1177/0361684316655757

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and Science Attitudes and Achievement at the Intersection of Gender and Ethnicity. *Psychology of Women Quarterly*, *37*(3), 293–309. https://doi.org/10.1177/0361684313480694

Evans, A. B., Copping, K. E., Rowley, S. J., & Kurtz-Costes, B. (2011). Academic self-concept in Black adolescents: Do race and gender stereotypes matter? Self and Identity, 10(2), 263-277. https://doi.org/10.1080/15298868.2010.485358

Feliciano, C. (2012). The female educational advantage among adolescent children of immigrants. Youth & Society, 44(3), 431-449.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519

Graham, S., & Morales-Chicas, J. (2015). The Ethnic Context and Attitudes toward 9th Grade Math. *International Journal of Educational Psychology*, *4*(1), 1–32.

Grund, S., Lüdtke, O., & Robitzsch, A. (2016). Pooling ANOVA results from multiply imputed datasets: A simulation study. Methodology, 12, 75-88. https://doi.org/10.1027/1614-2241/a000111

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of research in science teaching*, 39(9), 870-888.

Hedges, L. V., & Olkin, I. (1985). *Statistical Method for Meta-Analysis*. Academic Press. http://gbv.eblib.com/patron/FullRecord.aspx?p=1901162

Hedges, L. V., & Schauer, J. M. (2019). More Than One Replication Study Is Needed for Unambiguous Tests of Replication. *Journal of Educational and Behavioral Statistics*, *44*(5), 543–570. https://doi.org/10.3102/1076998619852953

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989x.3.4.486

Honey, M., Alberts, B., Bass, H., Castillo, C., Lee, O., Strutchens, M. M., Vermillion, L., & Rodriguez, F. (2020). *STEM education for the future: A vision report*.

Hsieh, T.-Y., Simpkins, S. D., & Eccles, J. S. (2021). Gender by racial/ethnic intersectionality in the patterns of Adolescents' math motivation and their math achievement and engagement. *Contemporary Educational Psychology*, *66*, 101974. https://doi.org/10.1016/j.cedpsych.2021.101974

Hyde, J. S. (2007). New Directions in the Study of Gender Similarities and Differences. *Current Directions in Psychological Science*, *16*(5), 259–263. https://doi.org/10.1111/j.1467-8721.2007.00516.x

Jiang, S., Simpkins, S. D., & Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Developmental Psychology*, *56*(11), 2137–2151. https://doi.org/10.1037/dev0001110

Jones, M. H., Audley-Piotrowski, S. R., & Kiefer, S. M. (2012). Relationships among adolescents' perceptions of friends' behaviors, academic self-concept, and math performance. *Journal of Educational Psychology*, *104*(1), 19–31. https://doi.org/10.1037/a0025596

Lazarides, R., Dicke, A.-L., Rubach, C., Oppermann, E., & Eccles, J. S. (2021). Motivational profiles across domains and academic choices within Eccles et al.'s situated expectancy-value theoretical framework. *Developmental Psychology*, *57*(11), 1893–1909. https://doi.org/10.1037/dev0001250

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. Review of Educational research, 86(2), 602-640.

Lee, J., Lee, H. J., Song, J., & Bong, M. (2021). Enhancing children's math motivation with a joint intervention on mindset and gender stereotypes. Learning and Instruction, 73, 101416.

Levine, S. C., & Pantoja, N. (2021). Development of children's math attitudes: Gender differences, key socializers, and intervention approaches. Developmental Review, 62, 100997.

Marsh, H. W., Parker, P. D., Guo, J., Basarkod, G., Niepel, C., & Van Zanden, B. (2021). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. Journal of Personality and Social Psychology, 121(1), 168.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, *81*(Pt 1), 59–77. https://doi.org/10.1348/000709910X503501

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. Journal for Research in Mathematics Education, 37(2), 129-150.

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The Development of Children's Gender-Science Stereotypes: A Meta-analysis of 5 Decades of U.S.

Draw-A-Scientist Studies. *Child Development*, *89*(6), 1943–1955. https://doi.org/10.1111/cdev.13039
Muenks, K., Wigfield, A., & Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Developmental Review*, *48*, 24–39. https://doi.org/10.1016/j.dr.2018.04.001
Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence*, *20*(2), 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x
National Center for Science and Engineering Statistics (NCSES) (2021). Women, minorities, and persons with disabilities in science and engineering. Washington, DC: National Science Foundation. Retrieved from https://ncses.nsf.gov/pubs/nsf21321/
National Academies of Sciences, Engineering, and Medicine (NASEM) (2020). Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors. Washington, DC: The National Academies Press. https://doi.org/10.17226/25585.
Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. American Economic Review, 106(5), 257-61. https://doi.org/10.1257/aer.p20161121
Parker, P. D., van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., & Noetel, M. (2020). The Intersection of Gender, Social Class, and Cultural Context: a Meta-Analysis. *Educational Psychology Review*, *32*(1), 197–228. https://doi.org/10.1007/s10648-019-09493-1
Parker, P. D., van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction*, *54*, 125–137. https://doi.org/10.1016/j.learninstruc.2017.09.002
Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology*, *95*(3), 589–603. https://doi.org/10.1037/0022-0663.95.3.589
Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796
Rosenbloom, S. R., & Way, N. (2004). Experiences of Discrimination among African American, Asian American, and Latino Adolescents in an Urban High School. *Youth & Society*, *35*(4), 420–451. https://doi.org/10.1177/0044118X03261479
Rosenzweig, E. Q., & Wigfield, A. (2016). STEM Motivation Interventions for Adolescents: A Promising Start, but Further to Go. *Educational Psychologist*, *51*(2), 146–163. https://doi.org/10.1080/00461520.2016.1154792
Rost, D. H., Sparfeldt, J. R., Dickhäuser, O., & Schilling, S. R. (2005). Dimensional comparisons in subject-specific academic self-concepts and achievements: A quasi-experimental approach. *Learning and Instruction*, *15*(6), 557–570. https://doi.org/10.1016/j.learninstruc.2005.08.003

Rowley, S. J., Kurtz-Costes, B., Mistry, R., & Feagans, L. (2007). Social status as a predictor of race and gender stereotypes in late childhood and early adolescence. Social Development, 16(1), 150-168. https://doi.org/10.1111/j.1467-9507.2007.00376.x

Safavian, N. (2019). What makes them persist? Expectancy-value beliefs and the math participation, performance, and preparedness of Hispanic youth. *AERA Open, 5*(3), 2332858419869342.

Scherrer, V., & Preckel, F. (2019). Development of Motivational Variables and Self-Esteem During the School Career: A Meta-Analysis of Longitudinal Studies. *Review of Educational Research*, *89*(2), 211–258. https://doi.org/10.3102/0034654318819127

Seo, E., Shen, Y., & Alfaro, E. C. (2019). Adolescents' Beliefs about Math Ability and Their Relations to STEM Career Attainment: Joint Consideration of Race/ethnicity and Gender. *Journal of Youth and Adolescence*, *48*(2), 306–325. https://doi.org/10.1007/s10964-018-0911-9

Simpkins, S. D.; Fredricks, J. A.; Eccles, J. S. (2015). The role of parents in the ontogeny of achievement-related motivation and behavioral choices. *Monographs of the Society for Research in Child Development*, *80*(2).

Skaalvik, E. M., & Rankin, R. J. (1994). Gender differences in mathematics and verbal achievement, self-perception and motivation. *The British Journal of Educational Psychology*, *64*(3), 419–428.

Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. Journal of Educational Psychology, 82(3), 546.

Skinner, O. D., Kurtz-Costes, B., Vuletich, H., Copping, K., & Rowley, S. J. (2021). Race differences in Black and White adolescents' academic gender stereotypes across middle and late adolescence. Cultural Diversity and Ethnic Minority Psychology, 27(3), 537-545. https://doi.org/10.1037/cdp0000384

Tobin, D. D., Menon, M., Menon, M., Spatta, B. C., Hodges, E. V. E., & Perry, D. G. (2010). The intrapsychics of gender: A model of self-socialization. *Psychological Review, 117*(2), 601-622. https://doi.org/10.1037/a0018936

Trytten, D. A., Lowe, A. W., & Walden, S. E. (2012). "Asians are Good at Math. What an Awful Stereotype" The Model Minority Stereotype's Impact on Asian American Engineering Students. *Journal of Engineering Education*, *101*(3), 439–468. https://doi.org/10.1002/j.2168-9830.2012.tb00057.x

Umarji, O., Dicke, A.-L., Safavian, N., Karabenick, S., & Eccles, J. (2021). Teachers Caring for Students and Students Caring for Math Motivation: The Development of Culturally and Linguistically Diverse Adolescents' Math. *Journal of School Psychology*, *84*, 32–48. https://doi.org/10.1016/j.jsp.2020.12.004

Urdan, T., & Bruchmann, K. (2018). Examining the Academic Motivation of a Diverse Student Population: A Consideration of Methodology. *Educational Psychologist*, *53*(2), 114–130. https://doi.org/10.1080/00461520.2018.1440234

Wang, M.-T., & Degol, J. L. (2017). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for

Practice, Policy, and Future Directions. *Educational Psychology Review*, *29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-x

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*(6), 1594–1611. https://doi.org/10.1037/a0027838

Wigfield, A., & Eccles, J. S. (2002). Students' Motivation During the Middle School Years. In J. Aronson (Ed.), *Improving Academic Achievement: Impact of Psychological Factors on Education* (pp. 159–184). Elsevier. https://doi.org/10.1016/B978-012064455-1/50011-7

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (pp. 657–700). John Wiley & Sons, Inc.. https://doi.org/10.1002/9781118963418.childpsy316

Wood, D., Kurtz-Costes, B., Rowley, S. J., Okeke-Adeyanju, N. (2010). Mothers' academic gender stereotypes and education-related beliefs about sons and daughters in African American families. Journal of Educational Psychology, 102(2), 521-530. https://doi.org/10.1037/a0018481

Zusho, A., & Kumar, R. (2018). Introduction to the Special Issue: Critical Reflections and Future Directions in the Study of Race, Ethnicity, and Motivation. *Educational Psychologist*, *53*(2), 61–63. https://doi.org/10.1080/00461520.2018.1432362

Tables

Table 1. *Descriptive statistics of gender differences in competence-related beliefs across grade levels and datasets*

			Males						*Females*					
Grade level	Datasets[a]	Scale range	*n*	*M*	*SD*	*95% CI*		POMP	*n*	*M*	*SD*	*95% CI*		POMP
9	HSLS[b]	1-4	3310	2.96	1.30	2.92	3.01	65.39	3390	2.84	1.36	2.84	2.93	62.81
	CAMP	1-5	1296	3.34	0.79	3.30	3.38	58.48	1425	3.15	0.79	3.11	3.19	53.83
	PSID-CDS	1-7	179	5.00	1.36	4.80	5.20	66.73	175	4.89	1.26	4.70	5.07	64.81
	CAB	1-7	52	4.93	1.25	4.58	5.27	65.50	59	4.80	1.25	4.48	5.12	63.30
10	CAMP	1-5	864	3.35	0.79	3.29	3.40	58.68	878	3.13	0.80	3.08	3.19	53.33
	PSID-CDS	1-7	178	4.78	1.14	4.62	4.95	63.07	149	4.67	1.37	4.45	4.89	61.21
	CAB	1-7	79	4.83	1.22	4.56	5.10	63.80	76	4.78	1.22	4.51	5.06	63.07
	MSALT	1-7	193	5.16	1.64	5.00	5.33	69.35	209	4.94	1.61	4.78	5.10	65.67
11	HSLS[b]	1-4	3980	2.72	1.24	2.69	2.76	57.49	3900	2.58	1.33	2.54	2.62	52.67
	CAMP	1-5	742	3.32	0.79	3.27	3.38	58.03	821	3.13	0.77	3.08	3.19	53.30
	PSID-CDS	1-7	157	4.82	1.47	4.59	5.05	63.67	166	4.56	1.32	4.36	4.76	59.33
	MADICS	1-7	348	5.14	1.29	5.00	5.28	69.07	342	4.81	1.30	4.68	4.95	63.57
	CAB	1-7	23	5.01	1.33	4.45	5.56	66.77	32	4.55	1.32	4.08	5.02	59.10
12	CAMP	1-5	278	3.37	0.77	3.28	3.46	59.18	236	3.13	0.77	3.03	3.23	53.20
	PSID-CDS	1-7	116	4.75	1.51	4.47	5.03	62.52	144	4.43	1.48	4.19	4.67	57.19
	CAB	1-7	47	4.54	1.38	4.14	4.94	59.00	62	4.31	1.38	3.96	4.65	55.10
	MSALT	1-7	192	4.85	1.51	4.67	5.02	64.10	195	4.60	1.56	4.43	4.78	60.02

Notes. [a] Order of datasets according to age from youngest (top) to oldest dataset (bottom); POMP = Percent of Maximum Possible score.

[b] Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Table 2. *Effects of gender differences, combined effects and results on the heterogeneity of effects in 9th and 10th grade*

	9th grade						*10th grade*					
	F	*df*	*p*	*g*	*S.E.*	*95 % CI*	*F*	*df*	*p*	*g*	*S.E.*	*95 % CI*
MSALT	--			--			3.59	1, 397	.06	.14	.10	-.06; .33
CAB	0.27	1, 106	.61	.10	.19	-.27; .48	0.05	1, 150	.82	.04	.16	-.27; .36
PSID-CDS	0.68	1, 343	.42	.08	.11	-.13; .29	0.63	1, 320	.43	.09	.11	-.13; .31
CAMP	38.08	1, 2716	<.001	.24	.04	.17; .32	32.22	1, 1736	<.001	.28	.05	.18; .37
HSLS[b]	24.99	1, 480	<.001	.09	.02	.04; .14	--					
Combined effect				.14[b]	.06	.04; .25				.22[a]	.04	.14; .29
Heterogeneity	Q = 11.78, df = 3, p = .01; I^2 = 72.93						Q = 4.75, df = 3, p = .19; I^2 = 36.86					

Notes. Order of datasets according to age from youngest (top) to oldest dataset (bottom); S.E. = standard error. CI = confidence interval; [a] = refers to the fixed effect model based on the non-significance of the Q-statistic, [b] =refers to the random effect model based on the significance of the Q-statistic; Q = tests fixed effect model against random effect model, i.e., null hypotheses is that effect sizes are similar across datasets, which corresponds the fixed effect model; I^2 = indicates the percentage of variance of real differences in effect sizes; reported results included students' achievement, family financial background, and parent education as covariates. [b] SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Table 3. *Effects of gender differences, combined effects and results on the heterogeneity of effects in 11th and 12th grade*

	11th grade						*12th grade*					
	F	*df*	*p*	*g*	*S.E.*	*95 % CI*	*F*	*df*	*p*	*g*	*S.E.*	*95 % CI*
MSALT	--						3.81	1, 382	.05	.16	.10	-.04; .36
CAB	1.56	1, 50	.22	.35	.28	-.19; .89	0.76	1, 104	.39	.17	.19	-.21; .55
MADICS	11.04	1, 685	.001	.26	.08	.11; .41	--					
PSID-CDS	2.75	1, 311	.10	.19	.11	-.03; .41	2.66	1, 248	.10	.21	.13	-.03; .46
CAMP	22.60	1, 1558	<.001	.24	.05	.14; .34	12.44	1, 509	<.001	.31	.09	.14; .49
HSLS[b]	28.72	1, 480	<.001	.11	.02	.07; .15	--					
Combined effect				.14[a]	.02	.10; .18				.23[a]	.06	.12; .34
Heterogeneity	Q = 9.05, df = 4, p = .06; I^2 = 55.79						Q = 1.40, df = 3, p = .71; I^2 = 0.00					

Notes. See Table 2.

Table 4. *Descriptive statistics of math competence-related beliefs differences by gender and race/ethnicity across datasets*

			Males						Females					
Grade level	Dataset[a]	Scale range	*N*	*M*	*SD*	95% CI		POMP	*N*	*M*	*SD*	95% CI		POMP
African-Americans														
9	HSLS[b]	1-4	370	2.98	0.58	2.92	3.04	65.87	320	2.97	0.79	2.88	3.06	65.67
	PSID-CDS	1-7	78	4.61	1.69	4.24	4.99	60.23	73	4.71	1.68	4.32	5.09	61.95
10	PSID-CDS	1-7	88	4.34	1.01	4.13	4.55	66.51	68	4.51	1.75	4.09	4.93	58.51
11	HSLS[b]	1-4	420	2.74	0.88	2.66	2.83	58.12	440	2.73	1.12	2.62	2.83	57.58
	PSID-CDS	1-7	82	4.82	1.47	4.50	5.14	63.66	74	4.67	1.33	4.37	4.98	61.24
	MADICS	1-7	231	5.14	1.26	4.97	5.30	68.95	211	4.80	1.26	4.63	4.97	63.35
12	PSID-CDS	1-7	47	4.68	1.44	4.26	5.09	61.27	68	4.91	1.59	4.53	5.30	65.22
Asian-Amerians														
9	HSLS[b]	1-4	260	3.20	0.33	3.15	3.24	73.18	290	3.02	0.43	2.97	3.07	67.31
	CAMP	1-5	157	3.50	0.78	3.38	3.62	62.53	158	3.29	0.77	3.17	3.41	57.33
10	CAMP	1-5	114	3.53	0.72	3.40	3.66	63.23	114	3.26	0.72	3.13	3.39	56.55
11	HSLS[b]	1-4	340	2.83	1.34	2.69	2.98	61.07	340	2.63	1.40	2.48	2.78	54.19
	CAMP	1-5	120	3.39	0.74	3.26	3.53	59.83	111	3.14	0.75	3.00	3.28	53.50
12	CAMP	1-5	65	3.42	0.74	3.23	3.60	60.38	44	2.96	0.74	2.74	3.18	49.03
European-Americans														
9	HSLS[b]	1-4	2130	2.97	0.78	2.94	3.01	65.64	2180	2.84	0.74	2.81	2.87	61.35

	CAMP	1-5	143	3.32	0.80	3.19	3.46	58.10	167	3.16	0.80	3.03	3.28	53.88
	PSID-CDS	1-7	102	5.05	1.11	4.83	5.27	67.48	101	4.96	1.05	4.76	5.17	66.03
10	CAMP	1-5	109	3.32	0.80	3.17	3.46	57.90	107	3.16	0.80	3.01	3.31	53.93
	PSID-CDS	1-7	90	4.90	0.98	4.69	5.10	64.94	81	4.72	1.12	4.48	4.97	62.08
11	HSLS [b]	1-4	2520	2.74	0.90	2.70	2.78	58.00	2440	2.56	1.00	2.52	2.60	52.15
	CAMP	1-5	97	3.21	0.84	3.05	3.38	55.33	112	3.02	0.84	2.86	3.17	50.48
	PSID-CDS	1-7	75	4.81	1.22	4.53	5.08	63.43	92	4.55	1.18	4.31	4.79	59.12
	MADICS	1-7	117	5.13	1.33	4.89	5.37	68.78	131	4.63	1.33	4.63	5.09	64.35
12	CAMP	1-5	27	3.33	0.83	3.02	3.64	58.33	23	2.64	0.83	2.30	2.97	40.90
	PSID-CDS	1-7	69	4.74	1.30	4.43	5.05	62.38	76	4.36	1.08	4.11	4.60	55.92
Latinx-Americans														
9	HSLS [b]	1-4	560	2.91	1.60	2.77	3.04	63.59	590	2.71	0.78	2.65	2.78	57.13
	CAMP	1-5	996	3.31	0.79	3.26	3.36	57.80	1100	3.14	0.80	3.09	3.18	53.40
10	CAMP	1-5	640	3.32	0.78	3.26	3.38	58.08	657	3.10	0.79	3.04	3.16	52.58
11	HSLS [b]	1-4	690	2.66	1.45	2.56	2.77	55.48	690	2.50	1.38	2.40	2.61	50.12
	CAMP	1-5	525	3.33	0.79	3.26	3.39	58.18	598	3.15	0.78	3.09	3.21	53.75
12	CAMP	1-5	186	3.36	0.74	3.36	3.47	59.08	169	3.23	0.74	3.12	3.34	55.73

Notes. [a] Order of datasets according to age from youngest (top) to oldest dataset (bottom); POMP = Percent of Maximum Possible score. [b] Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Table 5. *Effects of gender differences, combined effects and results on the heterogeneity of effects in 9^{th} and 10^{th} grade within ethnicities/races*

	9^{th} grade						10^{th} grade					
	F	*df*	*p*	*g*	*S.E.*	*95% CI*	*F*	*df*	*p*	*g*	*S.E.*	*95% CI*
African-Americans												
HSLS	0.01	1, 100	.92	.02	.08	-.13; .17	--					
PSID-CDS	0.12	1, 148	.73	-.06	.16	-.38; .36	0.50	1, 152	.48	-.12	.16	-.44; .19
Combined effect				.00[a]	.07	-.14; .14						
Heterogeneity	Q = 0.17, df = 1, p = .68, I^2 = 0.00						--					
Asian-Americans												
HSLS	27.59	1, 50	<.001	.47	.09	.30; .64	--					
CAMP	5.65	1, 310	.02	.27	.11	.05; .49	7.86	1, 223	<.001	.38	.13	.11; .64
Combined effect				.39[a]	.07	.26; .53						
Heterogeneity	Q = 1.86, df = 1, p =.17, I^2 = 46.37						--					
European-Americans												
HSLS	31.27	1, 400	<.001	.17	.03	.11; .23	--					
CAMP	3.33	1, 305	.07	.20	.11	-.02; .42	2.13	1, 211	.15	.20	.14	-.07; .47
PSID-CDS	0.32	1, 194	.57	.08	.14	-.19; .36	1.14	1, 167	.17	.15	.15	-.13; .47
Combined effect				.17[a]	.03	.11; .23				.19[a]	.10	-.01; .39
Heterogeneity	Q = .45, df = 2, p = .80, I^2 = 0.00						Q = 0.02, df = 1, p = .89, I^2 = 0.00					
Latinx-Americans												
HSLS	6.14	1, 150	.01	.16	.06	.05; .28	--					
CAMP	26.63	1, 2091	<.001	.21	.04	.13; .30	25.04	1, 1292	<.001	.28	.06	.17; .39
Combined effect				.20[a]	.04	.13; .26						

Heterogeneity	Q =.52 , df = 1, p = .47, I^2 = 0.00

Notes. See Table 2.

Table 6. *Effects of gender differences, combined effects and results on the heterogeneity of effects in 11th and 12th grade within race/ethnicity*

	11th grade						12th grade					
	F	*df*	*p*	*g*	*S.E.*	*95% CI*	*F*	*df*	*p*	*g*	*S.E.*	*95% CI*
African-Americans												
MADICS	7.44	1, 437	.01	.27	.10	.08; .46	--					
PSID-CDS	0.39	1, 148	.53	.11	.16	-.21; .42	0.61	1, 107	.44	-.15	.19	-.52; .22
HSLS	0.07	1, 100	<.001	.01	.07	-.12; .14	--			--		
Combined effect				.10[b]	.12	-.00; .20				--		
Heterogeneity		Q = 4.89, df = 2, p = .09, I^2 = 59.13										
Asian-Americans												
HSLS	11.51	1, 80	<.001	.15	.08	-.01; .30	--					
CAMP	6.58	1, 226	.01	.34	.13	.08; .60	9.83	1, 104	<.001	.62	.20	.23; 1.01
Combined effect				.19[a]	.07	.06; .32						
Heterogeneity	Q = 1.53, df = 1, p = .22, I^2 = 34.73											
European-Americans												
HSLS	48.70	1, 410	<.001	.19	.03	.13; .25	--					
CAMP	2.80	1, 204	.10	.23	.14	-.05; .50	8.60	1, 45	.01	.83	.30	.25; 1.41
PSID-CDS	1.82	1, 162	.07	.22	.16	-.09; .52	3.36	1, 140	.07	.32	.17	-.01; .65
MADICS	2.46	1, 243	.12	.38	.13	.12; .63	--					
Combined effect				.20[a]	.03	.15; .25				.44[a]	.15	.16;.73
Heterogeneity		Q = 2.07, df = 3, p = .56, I^2 = 0.00							Q = 2.27, df = 1, p = .13, I^2 = 55.95			
Latinx-Americans												
HSLS	4.60	1, 180	.03	.11	.05	.01; .22	--					

CAMP	14.67	1, 1118	<.001	.23	.06	.11; .35	2.91	1, 350	.09	.18	.11	-.03; .38
Combined effect				.17[a]	.04	.09; .24						
Heterogeneity		Q = 2.08, df = 1 , p = .15, I^2 = 52.03										

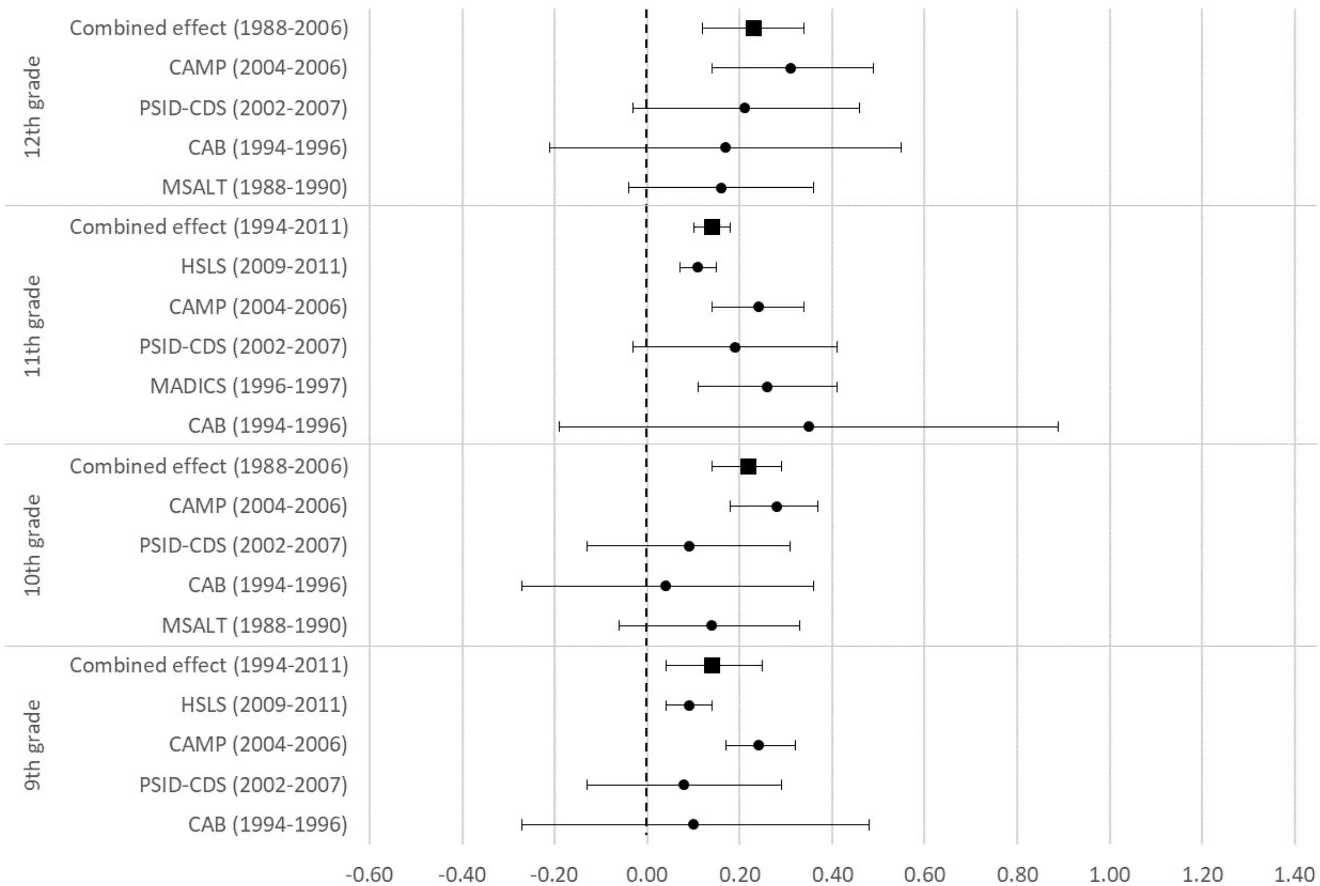
Notes. See Table 2.

Figures

Figure 1. *Used datasets by survey years and grade level that were used to investigate …*

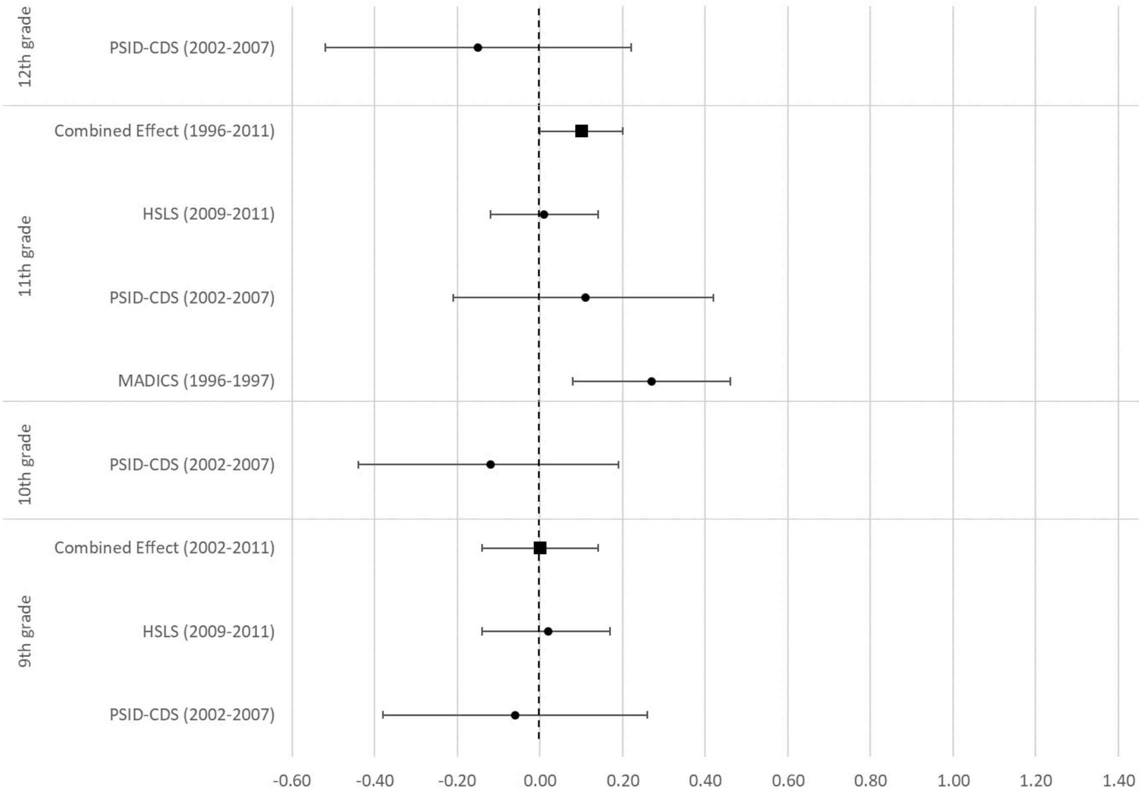
	9th grade	*10th grade*	*11th grade*	*12th grade*
… gender differences across all datasets				
1988-1990		MSALT		MSALT
1994-1996	CAB	CAB	CAB	CAB
1996-1997			MADICS	
2002-2007	PSID-CDS	PSID-CDS	PSID-CDS	PSID-CDS
2004-2006	CAMP	CAMP	CAMP	CAMP
2009-2011	HSLS		HSLS	
… gender differences within European Americans				
1996-1997			MADICS	
2002-2007	PSID-CDS	PSID-CDS	PSID-CDS	PSID-CDS
2004-2006	CAMP	CAMP	CAMP	CAMP
2009-2011	HSLS		HSLS	
… gender differences within Asian Americans				
2004-2006	CAMP	CAMP	CAMP	CAMP
2009-2011	HSLS		HSLS	
… gender differences within Latinx Americans				
2004-2006	CAMP	CAMP	CAMP	CAMP
2009-2011	HSLS		HSLS	
… gender differences within African Americans				
1996-1997			MADICS	
2002-2007	PSID-CDS	PSID-CDS	PSID-CDS	PSID-CDS
2009-2011	HSLS		HSLS	

Figure 2. *Gender effects and grade-specific weighted effect sizes (combined effects) from 9th to 12th grade*



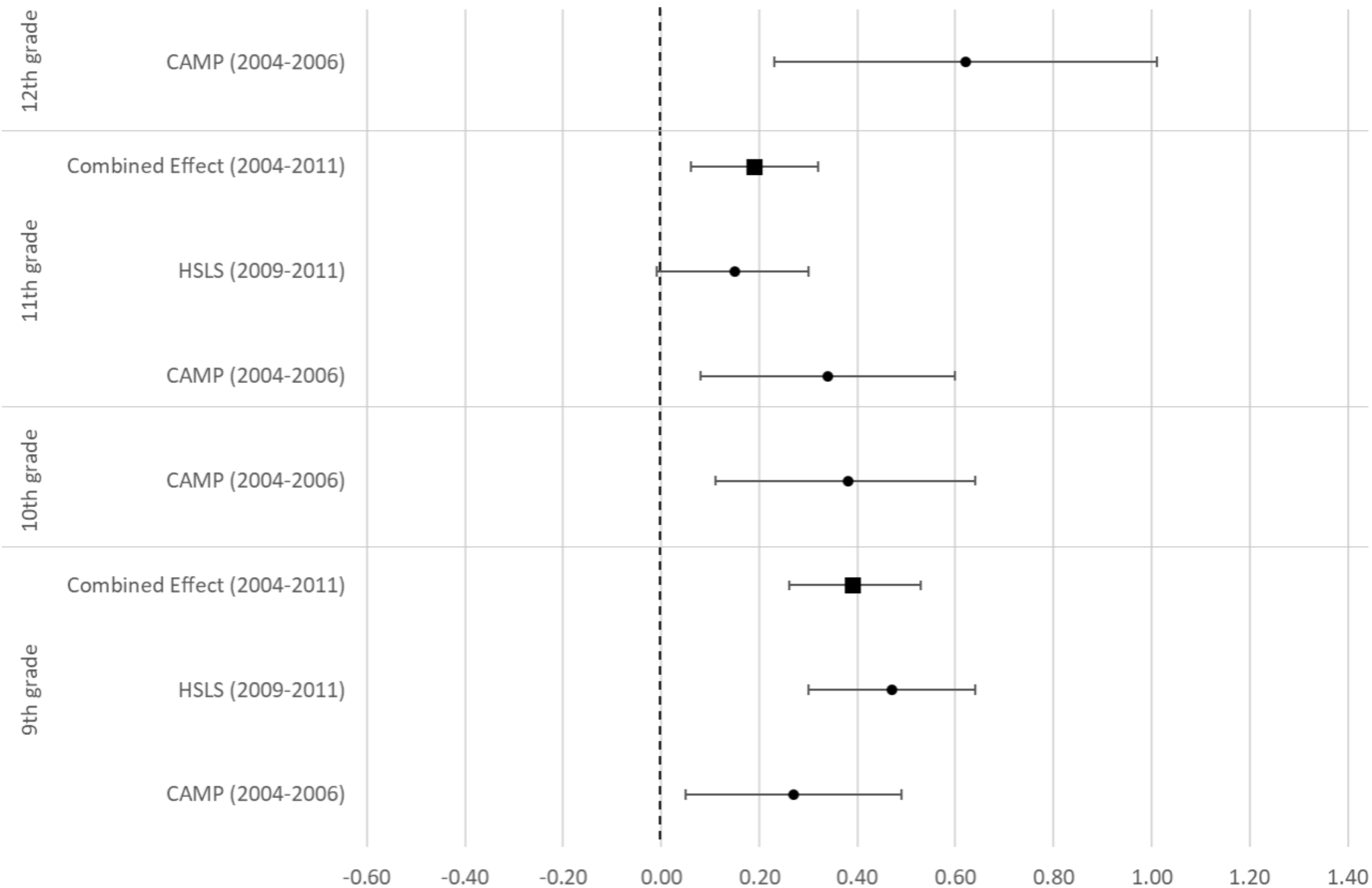
Notes. Reported results included student achievement, family financial background, and parent education as covariates; *Source HSLS:* U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Figure 3. *Gender effects and grade-specific weighted effect sizes (combined effects) for African-Americans from 9th to 12th grade*



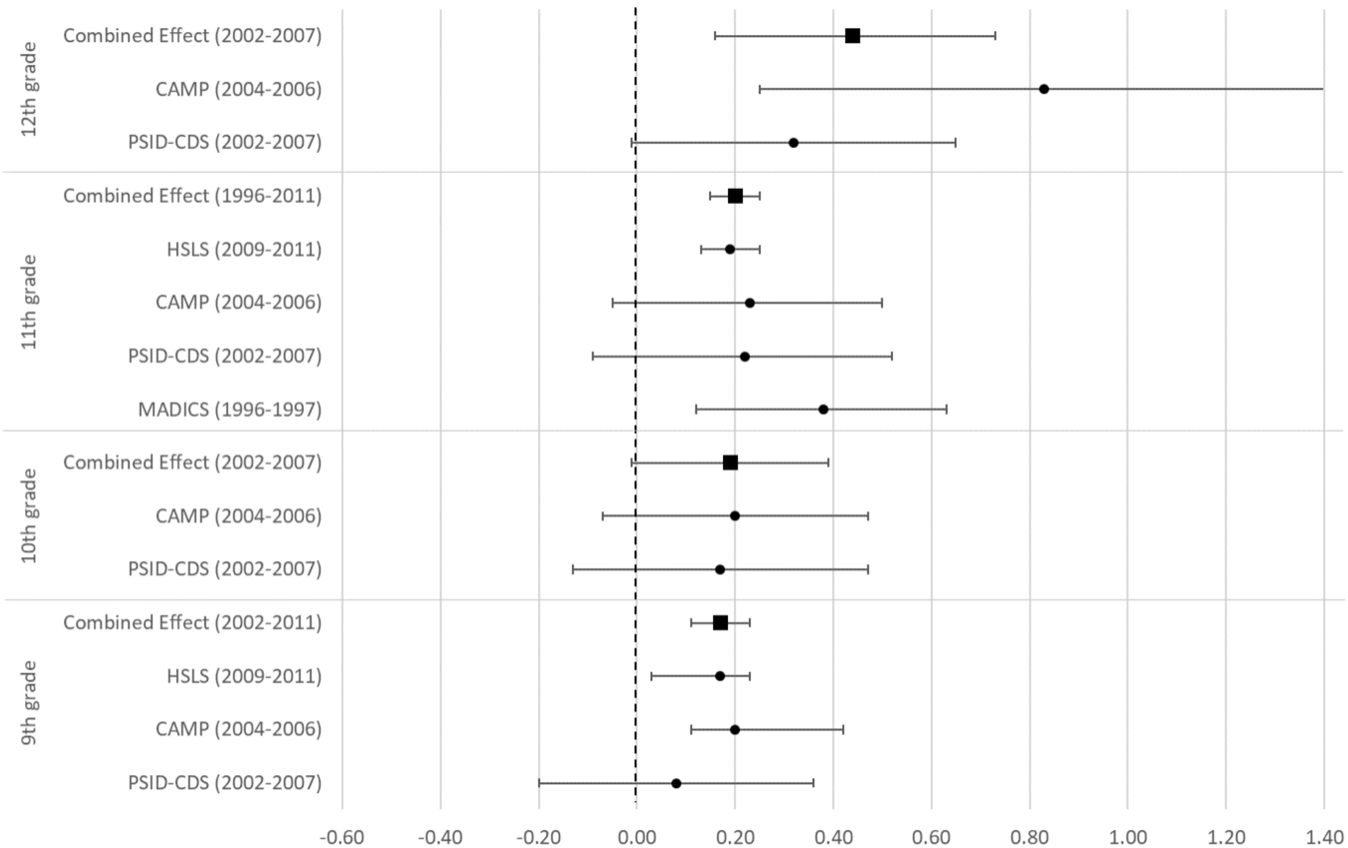
Notes. Reported results included student achievement, family financial background, and parent education as covariates; *Source HSLS:* U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Figure 4. *Gender effects and grade-specific weighted effect sizes (combined effects) for Asian-Americans from 9th to 12th grade*



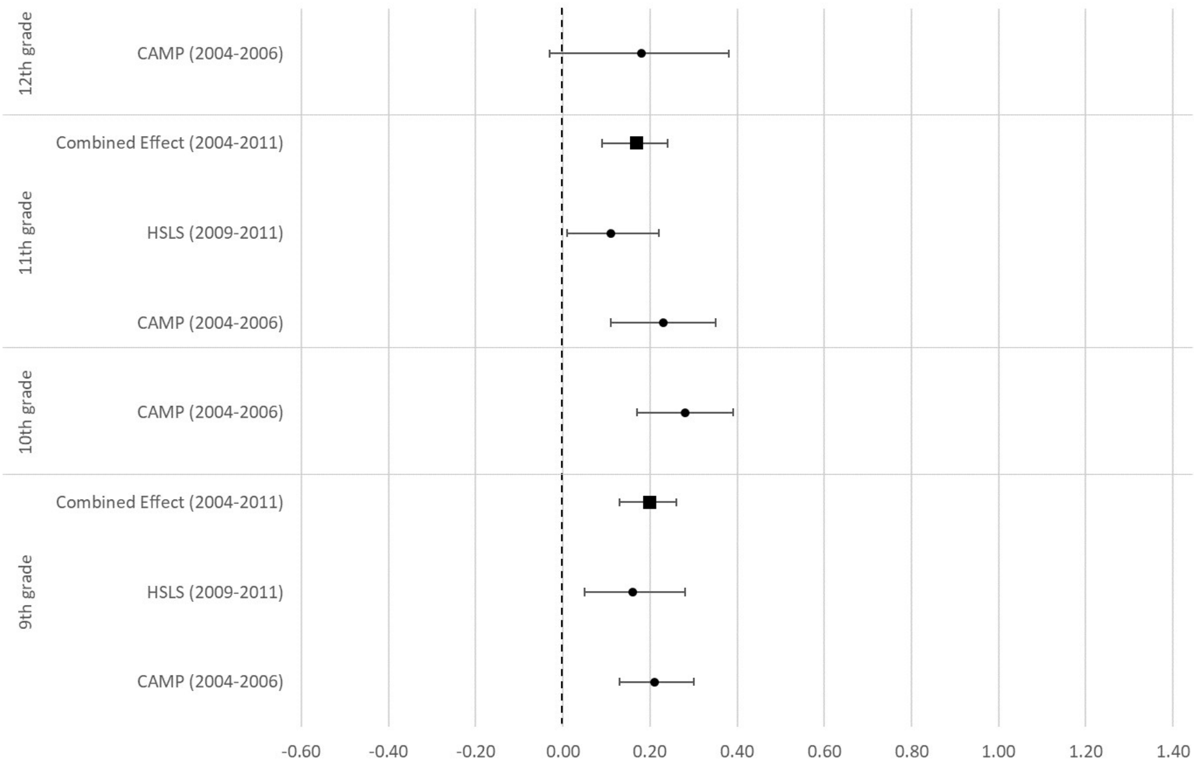
Notes. Reported results included student achievement, family financial background, and parent education as covariates; Source HSLS: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Figure 5. *Gender effects and grade-specific weighted effect sizes (combined effects) for European-Americans from 9th to 12th grade*



Notes. Reported results included student achievement, family financial background, and parent education as covariates; Source HSLS: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.

Figure 6. *Gender effects and grade-specific weighted effect sizes (combined effects) for Latinx-Americans from 9th to 12th grade*



Notes. Reported results included student achievement, family financial background, and parent education as covariates; Source HSLS: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year and First Year Follow-Up.