JOURNAL OF
Research on Adolescence

Society for Research on Adolescence
www.s-r-a.org

JOURNAL OF RESEARCH ON ADOLESCENCE, 29(2), 449–465

Threats and Supports to Female Students' Math Beliefs and Achievement

Sarah E. McKellar, Aixa D. Marchand, Matthew A. Diemer, and Oksana Malanchuk
University of Michigan

Jacquelynne S. Eccles
University of California, Irvine

This study examines how student perceptions of teacher practices contribute to female high school students' math beliefs and achievement. Guided by the expectancy–value framework, we hypothesized that students' motivation beliefs and achievement outcomes in mathematics are fostered by teachers' emphasis on the relevance of mathematics and constrained by gender-based differential treatment. To examine these questions, structural equation modeling was applied to a longitudinal panel of 518 female students from the Maryland Adolescent Development in Context Study. While controlling for prior achievement and race, gendered differential treatment was negatively associated with math beliefs and achievement, whereas relevant math instruction was positively associated with these outcomes. These findings suggest inroads that may foster positive math motivational beliefs and achievement among young women.

Students' motivation in math holds important implications for academic success and later career choices, especially related to science, technology, engineering, and mathematics (STEM) fields (e.g., Wang & Degol, 2013; Watt, 2006). In line with the expectancy–value framework, students are motivated to achieve in math when they believe they are capable (i.e., self-concept) and when they believe math is important (i.e., math importance, as part of task value) (e.g., Eccles, O'Neill, & Wigfield, 2005; Watt, Eccles, & Durik, 2006). School context matters for how these core motivational beliefs are formed and develop over time (Wigfield et al., 2015). More specifically, teachers are important socializing agents in the lives of adolescents, considering how much time adolescents spend in school (Eccles & Roeser, 2011; Roeser, Eccles, & Sameroff, 2000).

Teacher–student interactions may be especially important for female students' motivational beliefs (Eccles, 1994; Leaper & Brown, 2014; Spearman & Watt, 2013; Watt, 2006). Females are particularly sensitive to teacher feedback, internalizing classroom messages as diagnostic of their capabilities (Pomerantz, Altermatt, & Saxon, 2002). We focus on female students because they face more social identity threats to their math motivational beliefs than males—even though females may hold some advantages, such as higher levels of parental expectations and academic achievement than males (Brown & Leaper, 2010; Froiland & Davison, 2016; Leaper, Farkas, & Brown, 2012; Legewie & DiPrete, 2012; Pomerantz et al., 2002) .

For example, one of many threats outlined by prior work includes girls' perceptions that high school STEM classes were inhospitable to them due to their gender (Leaper et al., 2012). Leaper and Brown's (2014) review of discrimination in schools highlights how teachers may play a pernicious role in female students' declined self-concept, interest, and achievement in STEM fields. These social identity threats may partially account for findings that, beginning in early adolescence, female students are less likely to believe they are as competent in or to value math and science as much as male students (Fredricks & Eccles, 2002; Kurtz-Costes, Rowley, Harris-Britt, & Woods, 2008; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005).

On the other hand, females are likely to benefit from instructional strategies and interventions focused on the relevance of math and science (e.g., relevance interventions; Gaspard et al., 2015). Pedagogical emphases upon relevance may be one lever to foster female students' math beliefs and achievement. Yet, despite female students' apparent attunement to contextual inputs, few studies have examined gender-related social identity threats to motivational beliefs and achievement across middle

The third, fourth, and fifth authors were supported by grants from the National Science Foundation (#110878; #1231347), awarded to PIs Jacquelynne S. Eccles and Oksana Malanchuk. The Maryland Adolescent Development in Context Study (MADICS) was supported by grants from NICHD (#R01 HD33437) and the MacArthur Research Network on Successful Adolescent Development in High Risk Settings to J. S. Eccles and A. Sameroff. We thank those who support of the MADICS project and the reviewers for their helpful suggestions.

Requests for reprints should be sent to Sarah E. McKellar, Combined Program in Education and Psychology, University of Michigan, 610 East University Avenue, Ann Arbor, MI. E-mail: smckella@umich.edu

DOI: 10.1111/jora.12384

school and high school, and rarely have threats been examined in conjunction with supports of math motivational beliefs and achievement.

To better understand these issues, we examined how teachers' relevant math instruction (i.e., emphasizing the relevance of math for everyday life and making math interesting) and differential treatment based on gender relate to female students' math motivational beliefs and achievement over time. Declines in motivational beliefs during adolescence, particularly for females, have lasting consequences for the life course and may help to explain the dearth of qualified women in STEM positions (e.g., math-intensive fields such as computer science, physics, engineering; National Science Foundation [NSF], National Center for Science and Engineering Statistics [NCSES], 2015). The underrepresentation of women in STEM positions has historically been explained by lower achievement than men. Yet, the closing of the gender achievement gap in the past few decades has brought attention to other explanations. Thus, contextual factors and motivational beliefs, in addition to achievement, are important to understand to support female students' persistence in STEM fields (Riegle-Crumb, King, Grodsky, & Muller, 2012).

A longitudinal study by Legewie and DiPrete (2012) found evidence that high school contexts were related to female students' persistence in STEM fields. Motivational and contextual factors are important in understanding the long-term achievement of young women in STEM domains (Maltese & Tai, 2011). We focus on one specific academic domain (i.e., math) over time to gain a better understanding of how experiences in schools promote or threaten female students' math motivational beliefs and achievement. In line with the expectancy–value framework, students are motivated to achieve in math when they believe they are capable (i.e., self-concept) and when they believe math is important (i.e., math importance, as part of task value) (e.g., Eccles et al., 2005; Watt et al., 2006). In this study, we look at how female student perceptions of two contextual factors—gendered differential treatment and relevant math instruction—are related to two math motivational beliefs—self-concept and importance, as well as achievement. Furthermore, we look at these relationships for female students across middle school and high school.

Theoretical Framework: Expectancy–Value Model

We used the expectancy–value framework to investigate predictors of female students' math motivational beliefs and achievement (Eccles, et al., 2005). Within this model, students' expectancies of success and the value they have toward a domain or task predict student outcomes—namely, motivation and achievement. Therefore, two constructs central to this framework—self-concept of ability and importance value—hold strong implications for students' general achievement, as well as math academic success and STEM career-related choices (Eccles, 1994; Musu-Gillette, Wigfield, Harring, & Eccles, 2015; Simpkins, Davis-Kean, & Eccles, 2006; Wang & Degol, 2013). Eccles's (1983) expectancy-value model states that, while expectancies and self-concept of math ability are theoretically distinct, these constructs are empirically indistinguishable. Moreover, self-concept of math ability and task value become more positively related to one another as students move from early childhood into adolescence (Wigfield et al., 1997).

Within this model, self-concept of math ability, importance of math, and achievement are largely predicted by how students perceive the behaviors of influential adults, called "socializers" in the framework (Eccles, 1994, 2005). These theoretical predictors align with student perceptions of relevant math instruction and gendered differential treatment by teachers in this study. The expectancy–value framework aligns well to our research questions, as Wigfield et al. (2015) outline how achievement-related decisions are greatly influenced by cultural norms and socializing agents. Thus, this study investigates how motivational beliefs are related to teacher–student interactions as part of the classroom context, stressing the importance of context in academic beliefs and achievement (Eccles, 2005).

Teachers have the potential to transmit, reinforce, or challenge societal beliefs and stereotypes (e.g., girls are good at math due to "effort" and boys are good at math due to their "ability"; Fennema, Peterson, Carpenter, & Lubinski, 1990, p. 64; Gunderson, Ramirez, Levine, & Beilock, 2012). Although more egalitarian views of math ability by gender are becoming more prevalent (Leaper & Brown, 2014), research in the last decade has still identified negative stereotypes about females' abilities in math (Beilock, Gunderson, Ramirez, & Levine, 2010; Riegle-Crumb & Humphries, 2012). How students perceive their teachers' beliefs (e.g., that math is relevant) and behavior (e.g., gendered differential treatment) may play a significant role in self-concept, math importance value, and achievement.

School and Classroom Contexts

Gendered differential treatment by teachers. In the past few decades, observations and student reports of their perceptions have found that teachers interact differently with female students than with male students (e.g., being called on less, or thought of as less smart because of gender; Becker, 1981; Brown & Bigler, 2004, 2005; Spearman & Watt, 2013) . In studies of students in middle and high school contexts, over half of female students have experienced academic sexism—discouraging comments about girls' abilities in math, science, or computers (Brown & Bigler, 2004; Brown & Leaper, 2010; Leaper & Brown, 2008). Moreover, perceived differential treatment is highest in societal stereotyped fields (e.g., "boys have math and science abilities" and "girls have reading abilities" (Brown & Bigler, 2004, 2005). Negative stereotypes still persist, even though gender stereotypes in math have largely been overtaken by overtly egalitarian views of gender ability (Leaper & Brown, 2014). Studies have found that elementary school and high school teachers still perceive math as more difficult for girls, even when girls performed at the same level as boys (Cimpian, Lubienski, Timmer, Makowski, & Miller, 2016; Riegle-Crumb & Humphries, 2012). Thus, teacher biases, beliefs, and behaviors continue to play an important role in female student motivational beliefs, particularly in STEM subjects (Gunderson et al., 2012; Leaper & Brown, 2008; Riegle-Crumb & Humphries, 2012).

Over time, female students may internalize this gender-based (or gendered) differential treatment from teachers, which may undermine their own math beliefs and performance (e.g., Gunderson et al., 2012; Leaper & Brown, 2008; Wang & Degol, 2013). Perceived gender-based discrimination harms academic beliefs, values, and achievement (e.g., Brown & Bigler, 2004; Eccles & Roeser, 2011; Simpkins et al., 2006; Wang & Degol, 2013). For instance, when female students perceive discrimination from their teachers during adolescence, they believe they have lower science and math abilities than female students who do not perceive negative comments, irrespective of their science and math achievement and unlike their male peers (Brown & Leaper, 2010; Leaper et al., 2012).

For example, females and males assess their math competence differentially (e.g., girls underestimate their competence relative to boys, when controlling for achievement). This may lead to gender differences in their decisions to pursue careers in STEM fields (Cech, Rubineau, Silbey, & Seron, 2011; Correll, 2001). Moreover, Wang (2012) found females perceived that teachers believed girls had less "natural talent" and also had lower math motivational beliefs. Studies suggest that changes in females' cognitive development in middle school through high school allow for greater awareness of societal stereotypes, and this awareness may be linked with higher sensitivity to threats of discrimination (Brown & Bigler, 2004; Killen, Lee-Kim, McGlothlin, Stangor, & Helwig, 2002; Spearman & Watt, 2013). Thus, perceived gendered differential treatment introduces threats to female students' motivational beliefs relative to math.

Relevant math instruction. Instructional strategies aimed at promoting interest and relevance of math content may contribute to students' math ability beliefs and math importance (Spearman & Watt, 2013; Urdan & Schoenfelder, 2006; Wigfield et al., 2015). Relevant math instruction is conceptualized as how interesting and useful students perceive their teacher's delivery of the content. This conceptual and empirical approach mirrors prior research (e.g., Diemer, Marchand, McKellar, & Malanchuk, 2016).

There have been few studies that measure the effects of teaching practices that are relevant to students' lives in relation to student motivation over time (Diemer et al., 2016; Wang, 2012). However, there have been promising "one-shot" relevance interventions (i.e., interventions that take place at one or a few select time points; Gaspard et al., 2015; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009). These interventions typically promote the relevance of an academic domain (e.g., writing prompts about the real-world application of math or science) to increase students' motivation and achievement. Teachers often create similar opportunities for students to connect the content to what is relevant or important in their lives in their classroom instruction. When classroom teachers regularly create engaging and relevant experiences for students—similar to the interventions mentioned above—it translates into students' perceiving that the overall classroom instruction is interesting and important (Hidi & Renninger, 2006). Therefore, we extend this work by understanding how relevant math instruction—a perceived contextual support—is associated with motivational beliefs.

Student Motivational Beliefs

Self-concept of math ability. Self-concepts of ability (SCMA) are defined as beliefs about one's capability in a specified domain ("I'm good at math" vs. "I am good at reading"; Davis-Kean et al., 2008; Eccles et al., 2005). As students progress through school, they are better able to judge their abilities in comparison with their peers (i.e., external comparisons), as well as how their abilities are in one domain compared to another domain (i.e., internal comparisons). Additionally, as students transition from elementary school through high school, self-concepts of ability generally decrease, in part due to these internal and external comparisons (Denissen, Zarrett, & Eccles, 2007; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Self-concepts are important because students with a high self-concept of math value that subject more (e.g., Jacobs et al., 2002) and also tend to perform better in math and enroll in STEM courses in the future (Denissen et al., 2007; Simpkins & Davis-Kean, 2005; Simpkins et al., 2006; Wang, 2012; Wang & Degol, 2013).

Math importance. Math importance value—which we refer to as math importance for perspicuity—represents an aspect of math task value. Broadly, task value encompasses beliefs about the extent to which a domain (e.g., math, reading) is useful for the future, meaningful to one's sense of self, interesting, or costly (Eccles, 2005; Simpkins & Davis-Kean, 2005; Simpkins et al., 2006; Watt et al., 2006). Utility value (e.g., importance for the future) and attainment value (importance for sense of self) have been combined and referred to as "importance value" (e.g., Fredricks & Eccles, 2002; Jacobs et al., 2002; Watt, 2006; Watt et al., 2012). Thus, math importance is a more general construct related to math utility and attainment value, and is distinguished from math intrinsic value, or enjoyment of a subject (Gaspard et al., 2015; Watt et al., 2006, 2012). When students view that math is important, it predicts future choices and achievement in math and other STEM fields (Simpkins et al., 2006; Watt et al., 2006). Considered in concert, self-concept of math ability and importance become more closely related as students move from early childhood into adolescence (Fredricks & Eccles, 2002).

THE CURRENT STUDY

This study investigates the relationships between female students' perceptions of differential treatment and relevant math instruction, their reports of self-concept of math ability, math importance, and math achievement longitudinally, starting in the 8th grade and following students into the 11th grade. Our hypothesized paths between constructs of interest are shown in Figure 1. The following paths are outlined from the left (MADICS Wave 3, 8th grade) to the right (MADICS Wave 4, 11th grade). For simplicity, we did not depict paths from our control variables—prior (7th grade) math achievement and race—in our figure. Race was statistically controlled, given racial differences in math beliefs and outcomes (Wang & Degol, 2013).

In accordance with teacher influence in the expectancy–value model, we hypothesized that relevant math instruction will predict female students' self-concept of math ability and math importance. Moreover, we hypothesized that teachers' gendered differential treatment will negatively predict self-concept of math ability and math importance (Simpkins & Davis-Kean, 2005; Watt, 2006). Lastly, self-concept of math ability and math importance are hypothesized to predict achievement (i.e., math grades in the 8th grade and state standardized test scores in the 9th grade), as established by prior literature (Wigfield et al., 2015). This examination is unique in the sense that research has identified *downstream* consequences of math motivational beliefs (e.g., value, achievement, course taking) (Eccles & Wang, 2016; Simpkins et al., 2006; Wang, 2012; Watt et al., 2006), but few studies have longitudinally looked at *upstream* teacher practices that may shape these math motivational beliefs (Wang, 2012).

METHOD

Data Source

This study analyzed data from the Maryland Adolescent Development in Context Study (MADICS; Eccles, 1997), a large-scale longitudinal data set that richly measures constructs highly congruent with the expectancy–value framework over time. Starting in the fall of 1991, a sample of 1,482 seventh graders was surveyed from all 23 public middle schools in Prince George's County, Maryland. This study uses data from Wave 3 in 1993 (end of students' 8th-grade year) and Wave 4 in 1996 (during 11th grade). Waves 3 and 4 richly measure youths' perceptions of differential treatment, their schools, and their academic beliefs and achievement.

At Wave 3, the MADICS sample included 1,065 adolescents, 48.5% of whom were female. Due to our substantive focus on the impacts of gender-related processes on math beliefs and achievement, we created a subgroup of 518 female participants.

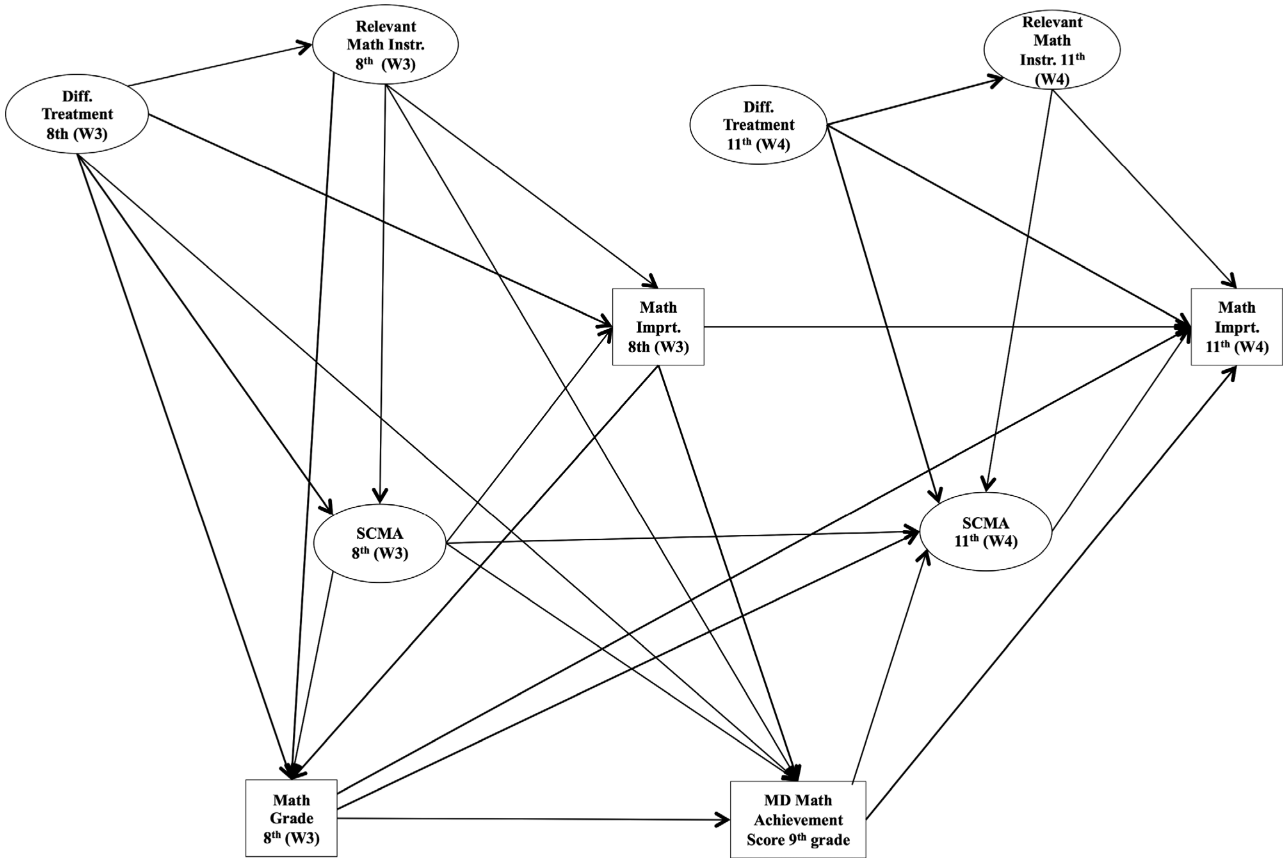

FIGURE 1 Hypothesized relations among constructs.

This female sample included youth who identified as African American ($n = 282$; 55%), White ($n = 176$; 34%), or as bi-racial or as a member of another racial/ethnic minority group ($n = 58$; 11%). Thirty percent of the parents of female participants reported that their highest level of education was high school or less, 19.4% reported that they had completed some college, 12.9% reported that they had a 4-year postsecondary degree, and 20.5% reported that they completed a graduate degree.

General Methodological Approach

This study uses structural equation modeling (SEM) to analyze these data, which simultaneously estimates hypothesized relations among numerous constructs while adjusting for measurement error. As part of this process, SEM precisely evaluates how well complex models fit the data (Kline, 2010). SEM is especially useful in secondary analyses, where all aspects of a latent construct may not be measured but available indicators can represent that construct. By accounting for measurement error, more precise tests of direct and indirect (i.e., mediated) relations are provided (Kline, 2010). Mplus version 7.4 (Muthén & Muthén, 2015) was used to conduct all SEM analyses.

Missingness in the data is depicted in Table 1. To make the missing at random assumption more tenable, these SEM analyses included auxiliary variables, which are likely predictors of missing data and attrition (e.g., they measure plausible missingness mechanisms), yet are not directly modeled as predictors or outcome variables. Instead, auxiliary variables correlate with each other while also predicting the residual terms of observed variables in the *saturated correlates* model (Enders, 2010). Previous inquiry identified a set of MADICS variables likely to be predictors of attrition and missing data (Diemer et al., 2016). These auxiliary variables were also employed in these analyses and include (1) standardized 5th-grade California Achievement Test math scores, (2) number of school absences in 7th grade, (3) self-reported likelihood participant will get involved with drugs later in life, (4) self-reported likelihood participant

TABLE 1
Descriptive Statistics for Variables and Observed Indicators

	8th grade Wave 3					11th grade Wave 4				
Variables/indicators and descriptions	*M*	*SD*	*Missing (%)*	*Alpha*	*IIC*	*M*	*SD*	*Missing (%)*	*Alpha*	*IIC*
Self-concept of math ability										
Scale 1–7, example 1 = *not at all good* to 7 = *very good*				0.82	0.6				0.85	0.66
v35170/v46038: How good are you in math?	5.04	1.65	1.2			4.6	1.71	17.8		
v35173/v46041: Compared to other kids your age how well do you do in math?	5.13	1.47	1.4			4.89	1.53	22.7		
v35181/v46049: Compared to other kids your age how well do you expect to do next year in math?	5.39	1.31	1.0			5.32	1.40	19.0		
Relevant math instruction										
Scale 1–5, all 1 = *almost never* to 5 = *always*				0.71	0.55				0.64	0.5
v35135/v46271: How often does your math teacher use examples that are interesting to you?	3.09	1.31	0.8			3.04	1.29	27.1		
v35137/v46273: How often do you learn things in math that help you with your everyday life?	3.17	1.25	0.8			2.62	1.25	27.1		
Math importance (1 item)										
Scale 1–7, all 1 = *much less important* to 7 = *much more important*				N/A	N/A				N/A	N/A
v35177/v46045: Compared to other kids your age, how important are each of the following activities to you: math?	5.11	1.50	1.0			4.82	1.57	22.7		
Perceived gendered differential treatment by teacher										
Scale 1–5, example l 1 = *never* to 5 = *more than six times*				0.79	0.44				0.83	0.53
v35304/v46350: How often do you feel like teachers call on you less often than they call on kids of the opposite sex?	1.58	0.87	1.6			1.37	0.74	25.8		
v35305/v46351: How often do you feel like you get disciplined more harshly by teachers than kids of the opposite sex?	1.44	0.83	1.6			1.26	0.65	25.8		
v35306/v46352: How often do you feel that teachers think you are less smart than kids of the opposite sex?	1.31	0.69	1.7			1.25	0.64	25.8		
v35309/v46353: How often have you felt that teachers/counselors discourage you from taking certain classes because of your sex?	1.24	0.60	2.9			1.20	0.56	25.8		
Math grade 7th grade (1 item) Scale 1–5	3.87	0.97	3.9							
Math grade 8th grade (1 item) Scale 1–5	3.70	0.94	11.4							
Maryland (MD) math achievement—9th grade (1 item) Range 273–421	357.25	28.61	19.6							

Note. IIC, mean inter-item correlation.

will get in trouble with the police later in life, and (5) parental educational attainment (a common measure of household socioeconomic status (SES); see Diemer, Mistry, Wadsworth, López, & Reimers, 2013). The auxiliary variables strategy also increases statistical power in that bias caused by missingness is attenuated—this increase in precision may reduce standard errors (Enders, 2010). All data were analyzed under full information maximum likelihood conditions. This strategy makes use of all existing data points in analyses (i.e., instead of deleting cases listwise or pairwise). Collectively, this is an efficient missing data strategy that maximizes effective sample size and statistical power for analyses (Enders, 2010).

Measures

Each measure is briefly reviewed below; further detail about each construct, observed items, the wording of and response options for each item, descriptive data, and internal consistency estimates are provided in Table 1. Because Cronbach's alpha is often a misleading estimate of internal consistency, in that it is downwardly biased when a measure consists of few items (Clark & Watson, 1995; DeVellis, 2003), mean inter-item correlations (which are not biased by the number of items in a measure) were also calculated. Generally, acceptable mean inter-item correlation values range from .15 to .50, with larger values reflecting higher levels of internal consistency (Clark & Watson, 1995). These measures have been used in prior work and have shown to be internally consistent, with IIC ranges between .49 and .61 and Cronbach's alpha between .79 and .88 (Diemer et al., 2016).

Perceived Gendered Differential Treatment by Teachers measures students' perceptions of differential treatment by teachers (and secondarily, school counselors), on the basis of gender (e.g., being called on less or thought of as less smart because of gender). These same items were also used in previous studies (Cogburn, Chavous, & Griffin, 2011; Roeser & Eccles, 1998; Roeser et al., 1998, 2000). This latent construct was developed from four Likert-type items that exhibited good internal consistency with this sample (see Table 1). It is important to note that perceived differential treatment in this data set is measured by students' reports of all teachers (as well as school counselors). Thus, this item measures general experiences of differential treatment rather than the extent to which the participants were experiencing gendered differential treatment from their math teachers, specifically. While it is plausible that gender-related discrimination in other school subjects and settings "spills over" into the mathematics classroom, this issue is returned to in the Limitations section of this paper.

Relevant Math Instruction represents student perceptions of how personally meaningful and relevant they believe their math curriculum to be (Diemer et al. 2016). Student perceptions of instruction, used here, are less likely to be influenced by social desirability than teachers' perceptions of their instruction (e.g., teachers might report more favorably on their own teaching practices than student reports of their teacher's practice), and are therefore likely more accurate measures (DeVellis, 2003). Two items measured relevant math instruction and evinced good internal consistency, as depicted in Table 1.

Self-Concept of Math Ability represents one's beliefs regarding math ability, and was measured by three Likert-type items that are consistent with previous measures of self-concept of academic ability (Denissen et al., 2007; Jacobs et al., 2002; Simpkins et al., 2006). As depicted in Table 1, self-concept of math ability exhibited good internal consistency with this population.

Math Importance refers to the subjective valuation of math (Eccles et al., 2005; Hentges & Wang, 2017; Watt, 2006). This one-item measure was modeled as an observed indicator in analyses, and as such was not included in the measurement model but only in the structural model analyses.

Math Achievement was measured by transcript grades and standardized math achievement scores. During 7th grade (MADICS Wave 1) and 8th grade (MADICS Wave 3) math achievement was measured by cumulative math grade point average for that academic year (on a 1–5 scale), obtained from school records. The Maryland Math Test, a state-level standardized achievement test, was administered to participants in 9th grade (between MADICS Wave 3 [8th grade] and Wave 4 [11th grade]) and was linked to the MADICS data set.

The Wave 1 (7th grade) math achievement measure preceded the Wave 3 (8th grade) and Wave 4 (11th grade) constructs of interest. Therefore, Wave 1 math achievement is modeled as a prior control variable in the structural model (see Figure 1). Modeling lagged achievement is a particularly strong strategy for addressing unobserved variables bias in educational research, in that prior achievement likely covaries with other observed and unobserved variables (Frank, 2000).

TABLE 2
Measurement Model: Factor Loadings for Latent Variables

Latent variable and indicators	*Standardized*	*SE*
Self-concept of math ability Wave 3		
v35170: How good are you in math?	.90*	.02
v35173: Compared to other kids your age how well do you do in math?	.85*	.02
v35181: Compared to other kids your age how well do you expect to do next year in math?	.60*	.04
Self-concept of math ability Wave 4		
v46038: How good are you in math?	.93*	.02
v46041: Compared to other kids your age how well do you do in math?	.90*	.03
v46049: Compared to other kids your age how well do you expect to do next year in math?	.63*	.04
Relevant math instruction Wave 3		
v35135: How often does your math teacher use examples that are interesting to you?	.77*	.04
v35137: How often do you learn things in math that help you with your everyday life?	.71*	.04
Relevant math instruction Wave 4		
v46271: How often does your math teacher use examples that are interesting to you?	.85*	.09
v46273: How often do you learn things in math that help you with your everyday life?	.56*	.07
Perceived gendered differential treatment by teacher Wave 3		
v35304: At school, how often do you feel like teachers call on you less often than they call on kids of the opposite sex?	.74*	.04
v35305: At school, how often do you feel like you get disciplined more harshly by teachers than kids of the opposite sex?	.84*	.03
v35306: At school, how often do you feel that teachers think you are less smart than kids of the opposite sex?	.78*	.04
v35309: How often have you felt that teachers/counselors discourage you from taking certain classes because of your sex?	.48*	.06
Perceived gendered differential treatment by teacher Wave 4		
v46350: At school, how often do you feel like teachers call on you less often than they call on kids of the opposite sex?	.81*	.03
v46351: At school, how often do you feel like you get disciplined more harshly by teachers than kids of the opposite sex?	.85*	.03
v46352: At school, how often do you feel that teachers think you are less smart than kids of the opposite sex?	.87*	.03
v46353: How often have you felt that teachers/counselors discourage you from taking certain classes because of your sex?	.71*	.05

Note. *Significant standardized estimates.

RESULTS

Measurement Model

By conducting a confirmatory factor analysis of student responses for both waves simultaneously, we were able to conclude that the 18 items significantly and strongly loaded onto the six latent constructs of interest (see Table 2). We established a good fitting measurement model, root mean square error of approximation (RMSEA) = 0.03, comparative fit index (CFI) = 0.98, Tucker-Lewis index (TLI) = 0.98, standardized root mean square residual (SRMR) = 0.04. Common sources of error variance are likely between items repeated across Waves 3 and 4, and therefore we estimated the correlated residual terms of repeated items in SEM analyses (Kline, 2010).

Measurement Invariance

Gender measurement invariance. Our research questions are centered on female students because of the gendered threats to mathematics success they face, as well as females' underrepresentation in particular STEM fields. Yet, male students may also perceive gendered threats to math achievement. We could not identify any inquiry regarding gendered differential treatment by teachers negatively affecting males' math beliefs and achievement. Nonetheless, the MADICS data set also contains males' responses about whether they feel they are differentially treated based on their gender, which could potentially allow for examinations of these processes for young men and for young women. Accordingly, we tested the measurement invariance of our differential treatment latent construct for males and for females. We were able to achieve configural and metric invariance, but could not establish scalar invariance. Males and females interpret the scaling of differential treatment items in distinct ways (see Table S2 in the online Supporting Information). We could not establish full measurement invariance, and accordingly we cannot compare these processes between men and women without being able to "rule out" any observed differences as caused by measurement differences between men and women (Kline, 2010). Because threats to male students' math beliefs and achievement are less compelling, both in terms of "broader impacts" and scholarship, and because we could not establish measurement invariance here, we focus on female students in all subsequent analyses. Analyses of the males-only subgroup are included in the online supplement to this article (see Figure S1 in the online Supporting Information).

Temporal measurement invariance. We tested whether latent variables measured at both waves (i.e., relevant math instruction, self-concept of math ability, and differential treatment) measured the same construct and in the same way over time—that these measures were temporally invariant. First, we established configural invariance, which determines if the "configuration" of items loaded onto respective latent constructs in the same way at Waves 3 and 4 (Kline, 2010; Schmitt & Kuljanin, 2008). Second, we used Wald tests to establish loading or metric invariance (i.e., that the magnitude of each item loading was invariant over time). To do so, the loadings of items repeated in each wave were constrained to be equal over time, and this parameter constraint evaluated with the Wald test. Items loading onto the self-concept of math ability and relevant math instruction constructs were invariant across Waves 3 and 4 (i.e., all Wald tests of parameters constrained to be equal were not statistically significant; see Table S1 in the online Supporting Information). However, the four differential treatment items were noninvariant over time. Accordingly, this means that gendered differential treatment by teachers means something different in 8th versus 11th grade. In some ways, this is not surprising, in that students have different teachers over time and approaches to discipline and instruction change from middle to high school. Yet, the invariance and noninvariance of focal constructs was considered in our interpretation of the results below. Additionally, *t*-tests were conducted to determine whether item means were significantly different across time (see Table S3 in the online Supporting Information). Of the perceived contextual factors items, one item measuring relevant math instruction and two items measuring gendered differential treatment items significantly decrease across time. Moreover, female motivational beliefs significantly declined; our math importance item and two of the self-concept of math ability items significantly decreased from middle school to high school.

Structural Model

The model depicted in Figure 1 was tested, while also controlling for race and prior (7th-grade) math achievement (For a table of direct and indirect paths, see Table S4 in the online Supporting Information.). This model had a good fit to the data (RMSEA = 0.05 CFI = 0.95, TLI = 0.93, SRMR = 0.08). The results for the completed structural model are detailed below and reported in Figure 2.

According to standard SEM notation, the effect size estimates are depicted by standardized coefficients (β). β coefficients obtained from SEM analyses are interpreted as they are in multiple regression analyses. These standardized estimates indicate how much change in the outcome variable is associated with a one-unit change in the predictor variable. Thus, coefficients <.10 are generally interpreted as small, coefficients between .30 and .50 medium, and coefficients >.50 large (Kline, 2010). Summarizing Wave 3 findings, gendered differential treatment by teachers in the 8th grade negatively related to student math importance ($\beta = -.09$) and math grade ($\beta = -.12$) within the same year. Differential treatment in the 8th grade also negatively predicted 9th-grade standardized achievement test scores ($\beta = -.17$).

In contrast, 8th-grade relevant math instruction was positively related to students' math importance ($\beta = .23$) and SCMA ($\beta = .54$) in the 8th grade. SCMA was also significantly related to 8th-grade math grades ($\beta = .29$) and math importance ($\beta = .54$) as well as scores on a state-level achievement test (Maryland Math Test) the following academic year ($\beta = .29$). Ninth-grade Maryland math test scores predicted 11th grade SCMA ($\beta = .32$).

Similarly, when examining significant paths for Wave 4, gendered differential treatment by teachers in the 11th-grade was negatively related to 11th-grade SCMA ($\beta = -.12$). Eleventh-grade relevant math instruction was positively related to 11th-grade SCMA ($\beta = .42$), and 11th-grade SCMA related to math importance ($\beta = .58$).

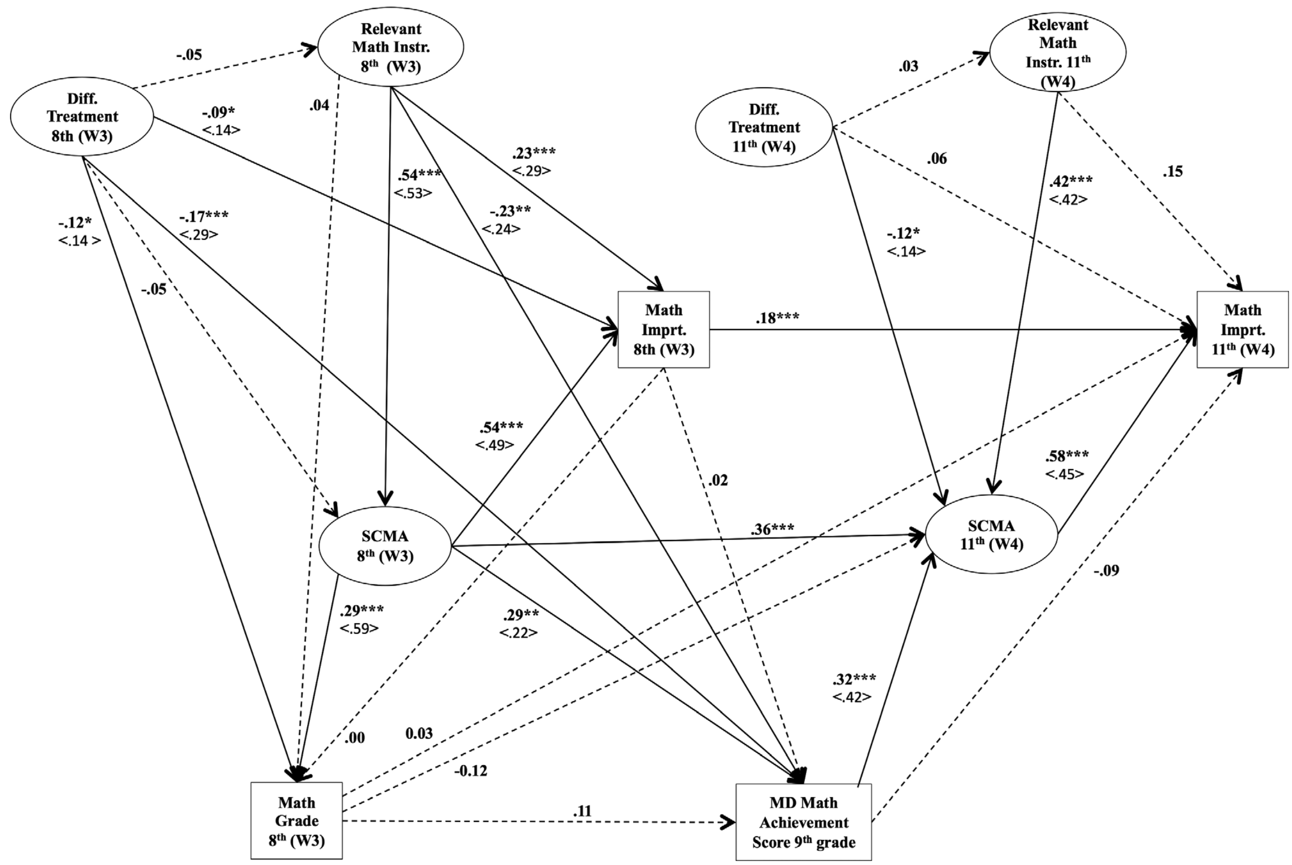

FIGURE 2 Structural model. Paths significant at the .05 level denoted with an asterisk * and a solid line. Nonsignificant paths are denoted by a dashed line. Impact threshold for a confounding variable (ITCV) estimates presented <inside brackets>.

We examined several substantively informed *indirect effects*—the term in SEM parlance, yet not asserting a cause and effect relationship—yet, these tests of partial mediation (i.e., indirect effects) were not depicted in Figure 2, for clarity. Eighth-grade relevant math instruction had an indirect effect upon 8th-grade math importance, via 8th-grade self-concept of math ability (β = .29). Furthermore, self-concept of math ability in the 8th grade partially mediated the relationship between 8th-grade relevant instruction and self-concept of math ability in the 11th-grade (β = .19). Eleventh-grade relevant instruction had an indirect effect upon 11th-grade math importance, via 11th-grade self-concept of math ability (β = .24). Lastly, Maryland Math Achievement scores in the 9th grade partially mediated the relationship between 8th-grade gendered differential treatment and self-concept of math ability in the 11th grade (β = .06).

Testing reverse causality models. One affordance of SEM is the capacity to test and compare substantively plausible alternative models. One such plausible model is a reverse causality model, which alters the direction of selected structural paths to evaluate the extent to which one construct appears to "cause" another construct (A→B), or vice versa (B→A)—yet, does not yield causal inferences from observational data (Pearl, 2009). Reverse causality modeling in this study investigated whether self-concept of math ability is instead a predictor of relevant math instruction. These reversed paths were tested because previous research suggests a reciprocal nature between teacher instruction and student motivation (Skinner & Belmont, 1993). Therefore, we reversed the regression paths between teachers' relevant math instruction and adolescents' self-concept of math ability at both Waves 3 and 4. Aside from the select reversed paths, reverse causality models are otherwise identical to the model being compared (here, the model depicted in Figure 1); the fit of the original structural and reverse causality models are then compared (Pearl, 2009).

Surprisingly, this reverse causality model had a fit equivalent to the structural model depicted in

Figure 1 and reviewed above (CFI = 0.95, TLI = 0.94, RMSEA = 0.04, SRMR = 0.07). This comparable fit suggests that reciprocal causation between teacher's relevant math instruction and students' self-concepts of math ability cannot be ruled out. Thus, it is possible that students' perceptions of their math ability predict their perceptions of teacher practices and conversely, these teacher practices may also promote students' perceptions of math ability. While reverse causality modeling does provide some evidence about the temporal ordering of constructs, causality cannot be ascertained. When the model fit does not favor the original or reverse causality model, theoretical considerations take on more importance. According to our guiding framework—expectancy–value theory—teacher practices are identified as preceding students' self-concept of math ability (Eccles et al., 2005). Altering the directions of regressions in this reverse causality model also would entail changing several pathways and be less consistent with current theory. Accordingly, we determined that the structural model shown in Figure 1 is the most substantively plausible and thus is our final model. Nevertheless, in the Discussion section, we discuss potential reciprocal causation.

Estimating the robustness of inferences. The impact threshold for a confounding variable (ITCV) is an estimate of how large an omitted variable would need to correlate with both a predictor and outcome variable to invalidate their statistically significant relationship (Frank, 2000). The ITCV coefficient estimates how robust a statistical inference is against unobserved variables bias, yet does not yield a causal estimate of the relation between a predictor and outcome variable.

Here, ITCV values were estimated for theoretically important relationships and are depicted in brackets in Figure 2 (for space reasons, only selected ITCV values are reviewed here). The ITCV calculation assumes that any unobserved confounder has no correlation with existing covariates, and as such is a conservative estimate. Because any unobserved confounder likely has some relation to observed covariates (in this case, race and prior math achievement), the "true" value of the ITCV—or the value needed to invalidate an inference about two variables—would most likely be larger (Frank, 2000).

Some obtained ITCV values for key relationships, such as for the significant relationship between relevant math instruction and self-concept of math ability at 8th grade (.54) and at 11th grade (.42), are quite large. This suggests that these relationships are more robust against the threat of unobserved variables bias, and therefore we can have more confidence in those relationships. The smaller but not negligible ITCV estimate for the relationship between relevant math instruction in the 8th grade and standardized math tests scores (.24) suggests that an unobserved variable could potentially nullify this inference if it were to be correlated above .24 with both variables. The implication of this finding is addressed in the Discussion section below.

DISCUSSION

This study examined how female students' experiences in their classrooms relate to math beliefs and achievement. Beyond confirming established findings that females' math motivational beliefs relate to one another and positively predict achievement (Denissen et al., 2007; Jacobs et al., 2002; Watt, 2006), this work advances understanding in the field in two main ways. First, perceived gendered differential treatment from teachers is negatively associated with female students' motivational beliefs and math achievement. Second, relevant math instruction is connected to positive outcomes for young women's motivational beliefs and indirectly relates to math achievement. Overall, this supports our hypotheses about how student perceptions of threats and supports are linked with math motivational beliefs and achievement (while controlling for prior math achievement and race).

A strength of this study was the ability to assess how student motivational beliefs and achievement change over time in relation to classroom contextual factors. We confirmed our hypotheses that female students' beliefs and experiences in middle school predict math motivational beliefs in high school (Wang, 2012). Specifically, we found links between female students' motivational beliefs in the 8th and 11th grades to their perceptions of whether their math instruction was relevant and whether teachers treated them differently than their male peers. How female students experience their school climate (relevant instruction and differential treatment) was particularly salient for their self-concept, especially since both reported experiences were linked with self-concept above and beyond prior motivational beliefs. This study adds to the growing literature that indicates that experiences during middle school have salient and potentially long-term impacts on future motivational beliefs. As early work by Pomerantz et al. (2002)

suggests, the middle school classroom context holds important implications, as social messages are salient for early adolescent students, especially girls (Eccles, 2009; Wang, 2012) .

Gendered Differential Treatment by Teachers

When examining students' perceptions of differential treatment by gender, we see these perceptions threaten female students' achievement and motivational beliefs. After experiencing discrimination in 8th grade, female students tended to have lower grades in the same year and lower test scores a year later. This work highlights that perceived threats in middle school may have long-term impacts on high school achievement. Since female students internalize academic feedback (e.g., grades) in ways that affect perceptions of their abilities and capabilities (Correll, 2001), a slight dip in math achievement could be perceived by a student as an indication that she is not good at math, despite the reality that she is doing just fine. This could be the case in our sample, as we see differential treatment in the 8th grade is linked to self-concept in the 11th grade through 9th-grade test scores.

We also see that perceptions of differential treatment are directly linked to motivational beliefs in middle school and in high school. Yet, this threat plays a different role in each context. In middle school, differential treatment is negatively associated with math importance; yet, in high school, differential treatment is negatively associated with students' sense of their math ability (self-concept). When interpreting these findings, it is important to consider the contextual and developmental changes that occur as students transition from middle school into high school. Theory suggests that youth perceptions of discrimination are linked to contextual factors, such as the ability to make a social comparison (e.g., demographics of student in the class), relevance of the discrimination to a stereotype, or perceived social support (Brown & Bigler, 2005). Few studies have compared these contextual factors that relate to gender discrimination between middle school and high school classrooms. Future work should look at differences in social support across context or social comparisons in light of course tracking in high school in order to further unpack how these processes function differently in various contexts.

Moreover, student developmental shifts affect their perceived discrimination. As Brown and Bigler (2005, p. 535) state, "By age 10, we expect that children's perceptions of discrimination will be fairly sophisticated and similar to that of adults, with the major exception of perceptions of societal or institutional discrimination." As students get older, their recognition of academic discrimination becomes more nuanced because of social influences and identity development (Leaper & Brown, 2008). These contextual and developmental changes may be linked to how the relationships between gendered differential treatment by teacher and motivational beliefs were different (i.e., noninvariant) between 8th grade and 11th grade. It may simply be that students change teachers and schools from 8th to 11th grade, and so therefore may experience gendered treatment in one school, and not the other. It may also be that developmental changes in students' capacity to discern discrimination lead them to be more able to perceive differential treatment in 11th grade. A greater understanding of how students across grades experience and interpret gendered differential treatment is an important direction for future research, especially given temporal noninvariance—suggesting student perceptions or interpretation of discrimination may change over time.

Relevant Math Instruction

Student perceptions that their math instruction was interesting and related to their everyday life are positively associated with motivational beliefs in the 8th grade and the 11th grade. In line with prior work and the expectancy–value framework, our study supports existing evidence that perceived instructional practices are related to math importance in middle school and self-concepts in high school (Fredricks & Eccles, 2002; Spearman & Watt, 2013; Wang, 2012).

Our results also revealed two surprising findings related to relevant math instruction: (1) the lack of a significant relationship between 8th-grade relevant instruction and 8th-grade math grades, and (2) the negative relationship between 8th-grade relevant instruction and 9th-grade standardized math achievement. It seems counterintuitive that relevant instruction does not directly and positively predict either achievement variable. However, relevant math instruction did have a positive indirect (or mediated) relationship with 8th-grade math grades and 9th-grade math achievement through 8th-grade self-concept of math ability. Thus, it may be that relevant math instruction fosters math achievement only insofar as it fosters math confidence first. Speculatively, "drill and kill" instructional strategies (which could not be examined with these data) may directly foster math achievement, particularly on standardized assessments—yet likely do not have the salutatory benefit of fostering mathematics self-concept of ability.

IMPLICATIONS

This study highlights the importance of links between perceived classroom climate, motivational beliefs, and achievement over time; these findings can inform the development of interventions. One possible way to support females' expectancies and values is to create relevant and egalitarian classroom environments in earlier grades and/or be more aware of students' perceptions of these situational factors (Brown & Bigler, 2004, 2016; Kurtz-Costes et al., 2008). As perceptions of academic discrimination appear to undermine the mathematics beliefs and achievement of young women, future work could do more to eliminate these experiences of discrimination in school settings—which also holds equity and justice implications.

Leaper and Brown (2014) argue that a willingness to acknowledge and confront issues of academic sexism may attenuate the STEM gender gap. As the expectancy–value theory outlines, teachers are important socializers of student motivational beliefs, especially for adolescent females (Eccles, 1994; Pomerantz et al., 2002). This study provides empirical evidence for how student self-concept and academic achievement can be damaged by influential adults in school contexts (Diemer et al., 2016; Eccles & Roeser, 2011). Teachers who behave in gender-biased ways (e.g., calling on females less often or thinking of them as being less smart than males) may negatively influence female students' beliefs, values, and long-term achievement. Conversely, teachers who utilize supports such as relevant instruction (e.g., interesting examples and connections to everyday life in math class) may bolster female students' math beliefs and achievement and thereby contribute to their persistence in STEM subjects.

LIMITATIONS AND FUTURE DIRECTIONS

While this study advances the field and our understanding of the importance of classroom contextual factors for STEM outcomes, further research may expand upon the current findings through the use of additional data and statistical approaches. One limitation is the inability to rule out reciprocal causality—that perceptions of teacher practices and students' math beliefs may be dynamic and mutually reinforcing processes—such that students with higher levels of math self-confidence, for example, may perceive their teachers as providing more relevant math instruction. This premise is supported by the results of the reverse causality model tested. Future inquiry, which ideally would contain multiple and closely spaced measurements, could better illuminate the intricacies of teacher practices and student beliefs.

Our current study was limited because math grades and standardized math achievement data were not collected for students in the 11th grade (MADICS Wave 4). Thus, we were not able to assess the extent to which our contextual constructs of interest in the 11th grade related to achievement in that same or later years. Nonetheless, we were able to study how our relationships of interest are linked to math achievement in 8th grade and 9th grade through examining both student grades and Maryland math test scores while simultaneously controlling for prior achievement (7th-grade math grades) and testing the robustness of inferences via the ITCV approach, which collectively bolsters confidence in these findings.

Moreover, students' math grade point average may be viewed as a less objective measure of student achievement, as prior work has identified teachers' biased expectations based on gender and race could influence grades (Ferguson, 2003; Riegle-Crumb & Humphries, 2012). On the other hand, standardized testing may also contain bias for females in subjects they are stereotyped to perform more poorly in. Therefore, we use both measures in order to understand how perceived contextual factors relate to achievement.

Future studies should examine how specific racial/ethnic groups experience gender-based differential treatment. MADICS does not provide sufficient sample sizes to examine these gendered processes with subsamples of female students from specific racial/ethnic groups (i.e., the White females sample size, $N = 129$, was too small to support these analyses). Therefore, these analyses controlled for race while focusing on gender. However, we recommend future work measure experiences of differential treatment accounting for the intersectionality of race/ethnicity and gender, namely how students experience distinct forms of discrimination, based on their multiple identities (e.g., Latino male, Asian female). For example, African American females may experience the classroom from the perspective of holding two oppressed identities. In fact, some work has specifically assessed perceptions of academic sexism from students with "double-minority status" (e.g., females in ethnically marginalized populations). Students who identify as having "double-minority status" are particularly vulnerable to threats against their competencies in STEM fields (Leaper et al., 2012). With our results highlighting how gendered differential treatment holds negative implications for females and prior work showing that teacher discrimination has negative implications for African American students' self-concept of math ability and achievement (Cogburn et al., 2011; Diemer et al., 2016), it is important to examine how these influences may have a unique and perhaps multiplicative effect on students who hold two marginalized identities.

Furthermore, a limitation of this study was our inability to control for school- or classroom-level effects. Information about the students' classrooms, or particularly math teachers, was not collected on this sample. Thus, we were unable to assess classroom-level predictors on individual student outcomes. In this data set, perceived differential treatment is measured by students' reports of all teachers rather than specific questions pertaining to students' math teachers. Due to this limitation, we were unable to determine the extent to which the participants were experiencing gender-based differential treatment from their math teacher specifically, another teacher, or all teachers. This information is particularly important for future studies to investigate, as gender-based discrimination from math teachers may have stronger implications for math beliefs and achievement. In addition to nesting the data within classrooms, we were unable to assess students' experience of differential treatment with a contemporary sample. Future work should examine whether more overt forms of gendered mistreatment persist, have increased, or have decreased. Regardless of how gender discrimination manifests in classrooms, over half of female students in a contemporary sample still report experiencing such treatment (Leaper & Brown, 2008). Thus, future work should build up this study to continue to identify threats to student motivational beliefs, particularly among students who experience multiple identities that are marginalized.

Lastly, we may better understand the threat of gendered differential treatment by examining different aspects of math importance (i.e., importance for the future vs. importance for identity) or other aspects of value. While our measure of math importance most closely aligns with utility value (the importance of math for future goals), the paths between constructs in our study may relate differently if we looked at different types of value, such as intrinsic, attainment, utility, or cost value (Eccles et al., 2005; Trautwein et al., 2012; Watt et al., 2006) . Thus, future directions should examine whether teacher–student interactions over an extended period of time may differentially relate to the various subdomains of value.

CONCLUSION

Instruction that promotes interest and brings attention to the importance of math positively relates to female students' confidence in their math ability (i.e., SCMA) in both the 8th and 11th grades—while also indirectly fostering math importance and achievement (as mediated by SCMA). These findings suggest inroads to foster math achievement and STEM success among young women. Even with narrowed gender gaps in certain STEM fields, Spearman and Watt (2013) point out how "the current social climate surrounding STEM subjects and workplaces often positions girls and women as less able than [boys and] men (even though there is a wealth of evidence to the contrary)" (p. 185). This study begins to respond to these concerns by exploring *upstream* factors that have been linked to *downstream* persistence in STEM fields—motivation and achievement in math throughout middle and high school. These findings advance our collective understanding of how experiences in schools promote or threaten female students' math motivational beliefs and achievement over time. Further studies examining school contextual factors and potential interventions in relation to students' math self-concepts and value are needed to address gaps among underrepresented groups in STEM fields.

REFERENCES

Becker, J. R. (1981). Differential treatment of females and males in mathematics classes. *Journal for Research in Mathematics Education*, *12*, 40–53. https://doi.org/10.2307/748657

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 1060–1063. https://doi.org/10.1073/pnas.0910967107

Brown, C. S., & Bigler, R. S. (2004). Children's perceptions of gender discrimination. *Developmental Psychology*, *40*, 714–726. https://doi.org/10.1037/0012-1649.40.5.714

Brown, C. S., & Bigler, R. S. (2005). Children's perceptions of discrimination: A developmental model. *Child Development*, *76*, 533–553. https://doi.org/10.1111/j.1467-8624.2005.00862.x

Brown, C. S., & Bigler, R. S. (2016). Feminist perspectives on gender development: Contributions to theory and practice. In T.-A. Roberts, N. Curtin, L. E. Duncan, & L. M. Cortina (Eds.), *Feminist perspectives on building a better psychological science of gender* (pp. 61–78). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-32141-7_5

Brown, C. S., & Leaper, C. (2010). Latina and European American girls' experiences with academic sexism and their self-concepts in mathematics and science during adolescence. *Sex Roles*, *63*, 860–870. https://doi.org/10.1007/s11199-010-9856-5

Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review*, *76*, 641–666. https://doi.org/10.1177/0003122411420815

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open*, *2*, 1–18. https://doi.org/10.1177/2332858416673617

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cogburn, C. D., Chavous, T. M., & Griffin, T. M. (2011). School-based racial and gender discrimination among African American adolescents: Exploring gender variation in frequency and implications for adjustment. *Race and Social Problems*, *3*, 25–37. https://doi.org/10.1007/s12552-011-9040-8

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments 1. *American Journal of Sociology*, *106*, 1691–1730. https://doi.org/10.1086/321299

Davis-Kean, P. E., Huesmann, L. R., Jager, J., Collins, W. A., Bates, J. E., & Lansford, J. E. (2008). Changes in the relation of self-efficacy beliefs and behaviors across development. *Child Development*, *79*, 1257–1269. https://doi.org/10.1111/j.1467-8624.2008.01187.x

Denissen, J. J., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*, 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.) Thousand Oaks, CA: Sage.

Diemer, M. A., Marchand, A. D., McKellar, S. E., & Malanchuk, O. (2016). Promotive and corrosive factors in African American students' math beliefs and achievement. *Journal of Youth and Adolescence*, *45*, 1208–1225. https://doi.org/10.1007/s10964-016-0439-9

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research: Social class measurement. *Analyses of Social Issues and Public Policy*, *13*, 77–113. https://doi.org/10.1111/asap.12001

Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75–146). San Francisco, CA: Freeman.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly*, *18*, 585–609. https://doi.org/10.1111/j.1471-6402.1994.tb01049.x

Eccles, J. S. (1997). Maryland Adolescent Development In Context Study (MADICS) [Measurement instrument]. Retrieved from http://www.rcgd.isr.umich.edu/pgc/

Eccles, J. S. (2005). Studying gender and ethnic differences in participation in math, physical science, and information technology. *New Directions for Child and Adolescent Development*, *2005*(110), 7–14. https://doi.org/10.1002/(ISSN)1534-8687

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*, 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., O'Neill, S. A., & Wigfield, A. (2005). Ability self-perceptions and subjective task values in adolescents and children. In K. A. Moore & L. H. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 237–249). Boston, MA: Springer. https://doi.org/10.1007/0-387-23823-9_15

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, *21*, 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Eccles, J. S., & Wang, M. T. (2016). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development*, *40*, 100–106. https://doi.org/10.1177/0165025415616201

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educational Studies in Mathematics*, *21*, 55–69. https://doi.org/10.1007/BF00311015

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black–White test score gap. *Urban Education*, *38*, 460–507. https://doi.org/10.1177/0042085903038004006

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods and Research*, *29*, 147–194. https://doi.org/10.1177/0049124100029002001

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

Froiland, J. M., & Davison, M. L. (2016). The longitudinal influences of peers, parents, motivation, and mathematics course-taking on high school math achievement. *Learning and Individual Differences*, *50*, 252–259. https://doi.org/10.1016/j.lindif.2016.07.012

Gaspard, H., Dicke, A. L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*, 1226–1240. https://doi.org/10.1037/dev0000028

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*, 153–166. https://doi.org/10.1007/s11199-011-9996-2

Hentges, R. F., & Wang, M. T. (2017). Gender differences in the developmental cascade from harsh parenting to educational attainment: An evolutionary perspective. *Child Development*. Advance online publication. https://doi.org/10.1111/cdev.12719.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*, 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*, 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412. https://doi.org/10.1126/science.1177067

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527. https://doi.org/10.1111/1467-8624.00421

Killen, M., Lee-Kim, J., McGlothlin, H., Stangor, C., & Helwig, C. C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the Society for Research in Child Development*, *67*, i–129.

Kline, R. B. (2010). *Principles and practice of structural equation modeling: Methodology in the social sciences* (3rd ed.) New York, NY: Guilford Press.

Kurtz-Costes, B. E., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender stereotypes about mathematics and science and self-perceptions of ability in late childhood and early adolescence. *Merrill-Palmer Quarterly*, *54*, 386–409. https://doi.org/10.1353/mpq.0.0001

Leaper, C., & Brown, C. S. (2008). Perceived experiences with sexism among adolescent girls. *Child Development*, *79*, 685–704. https://doi.org/10.1111/j.1467-8624.2008.01151.x

Leaper, C., & Brown, C. S. (2014). Sexism in schools. In L. Liben & R. Bigler (Eds.) *Advances in child development and behavior: The role of gender in educational contexts and outcomes* (Vol. 47, pp. 189–223). San Diego, CA: Elsevier. https://doi.org/10.1016/bs.acdb.2014.04.001

Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and English. *Journal of Youth and Adolescence*, *41*, 268–282. https://doi.org/10.1007/s10964-011-9693-z

Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, *77*, 463–485. https://doi.org/10.1177/0003122412440802

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, *95*, 877–907. https://doi.org/10.1002/sce.20441

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, *21*, 343–370. https://doi.org/10.1080/13803611.2015.1057161

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (Version 7.4). Los Angeles, CA: Muthén & Muthén.

National Science Foundation (NSF), National Center for Science and Engineering Statistics (NCSES). (2015). *Women, minorities, and persons with disabilities in science and engineering: 2015. Special report NSF 15-311*. Arlington, VA: National Science Foundation (NSF), National Center for Science and Engineering Statistics (NCSES). Retrieved from http://www.nsf.gov/statistics/wmpd/

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology*, *94*, 396–404. https://doi.org/10.1037/0022-0663.94.2.396

Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, *26*, 290–322. https://doi.org/10.1177/0891243211434614

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*, 1048–1073. https://doi.org/10.3102/0002831211435229

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (1998). Academic and emotional functioning in early adolescence: Longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology*, *10*, 321–352. https://doi.org/10.1017/S0954579498001631

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal*, *100*, 443–471. https://doi.org/10.1086/499650

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*, 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Simpkins, S. D., & Davis-Kean, P. E. (2005). The intersection between self-concepts and values: Links between beliefs and choices in high school. *New Directions for Child and Adolescent Development*, *2005*(110), 31–47. https://doi.org/10.1002/cd.148

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, *42*, 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581. https://doi.org/10.1037/0022-0663.85.4.571

Spearman, J., & Watt, H. M. G. (2013). Women's aspirations towards "STEM" careers. In W. Patton (Ed.), *Conceptualizing women's working lives: Moving the boundaries of discourse* (pp. 175–191). Rotterdam, The Netherlands: Sense Publishers. https://doi.org/10.1007/978-94-6209-209-9

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*, 763–777. https://doi.org/10.1037/a0027470

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology*, *44*, 331–349. https://doi.org/10.1016/j.jsp.2006.04.003

Wang, M. T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology*, *48*, 1643–1657. https://doi.org/10.1037/a0027247

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*, 304–340. https://doi.org/10.1016/j.dr.2013.08.001

Watt, H. M. (2006). The role of motivation in gendered educational and occupational trajectories related to math. *Educational Research and Evaluation*, *12*, 305–322. https://doi.org/10.1080/13803610600765562

Watt, H. M., Eccles, J. S., & Durik, A. M. (2006). The leaky mathematics pipeline for girls: A motivational analysis of high school enrolments in Australia and the USA. *Equal Opportunities International*, *25*, 642–659. https://doi.org/10.1108/02610150610719119

Watt, H. M., 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*, 1594–1611. https://doi.org/10.1037/a0027838

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (Vol. 3, pp. 1–44). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118963418.childpsy316

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*, 451–469. https://doi.org/10.1037/0022-0663.89.3.451

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Measurement Temporal Invariance Wald test Between 8th grade (W3) and 11th grade (W4)
Table S2. Multiple Group Measurement Invariance Test Between Female and Male Students
Table S3. *t*-Test Between 8th-Grade (W3) and 11th-Grade (W4) Items
Table S4. Overview of paths
Figure S1. Structural model for male students