

# The Consequences of the National Math and Science Performance Environment for Gender Differences in STEM Aspirations

Allison Mann, Thomas A. DiPrete

Columbia University

**Abstract:** Using the lens of expectation states theory, which we formalize in Bayesian terms, this article examines the influences of national performance and self-assessment contexts on gender differences in the rate of aspiring to science, technology, engineering, and mathematics (STEM) occupations. We demonstrate that girls hold themselves to a higher performance standard than do boys before forming STEM orientations, and this gender “standards gap” grows with the strength of a country’s performance environment. We also demonstrate that a repeatedly observed paradox in this literature—namely, that the STEM gender gap increases with a more strongly gender-egalitarian national culture—vanishes when the national performance culture is taken into account. Whereas other research has proposed theories to explain the apparent paradox as an empirical reality, we demonstrate that the empirical relationship is as expected; net of the performance environment, countries with a more gender-egalitarian culture have a smaller gender gap in STEM orientations. We also find, consistent with our theory, that the proportion of high-performing girls among STEM aspirants grows with the strength of the national performance environment even as the overall gender gap in STEM orientations grows because of offsetting behavior by students at the lower end of the performance distribution.

**Keywords:** education; gender inequalities; careers in science and technology; cross-cultural research

**Citation:** Mann, Allison, and Thomas A. DiPrete. 2016. “The Consequences of the National Math and Science Performance Environment for Gender Differences in STEM Aspirations.” *Sociological Science* 3: 568-603.

**Received:** February 22, 2016

**Accepted:** March 31, 2016

**Published:** July 12, 2016

**Editor(s):** Jesper Sørensen, Stephen Morgan

**DOI:** 10.15195/v3.a25

**Copyright:** © 2016 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

THE gender gap in science, technology, engineering, and mathematics (STEM) fields is a topic of much current interest. In addition to implications for the supply of qualified labor in STEM occupations and the consequent potential for economic growth, gender gaps in STEM fields raise questions about gender discrimination in education and labor markets. They contribute to gender inequality in earnings (Dey and Hill 2007) that persists despite the reversal of the gender gap in educational attainment (DiPrete and Buchmann 2013). Gender gaps in STEM fields also provide important insight into the continuing role of gender stereotypes in shaping the attitudes, expectations, and behaviors of girls and boys as they transition to the work force (Ridgeway 2011).

Studies on gender segregation in fields of study suggest that beliefs that certain characteristics are “essential” characteristics of gender, for example that boys are “naturally” better at math and science, perpetuate the sorting of boys and girls into different fields of study and occupations (Correll 2001; Eccles 1994). In the United States and most other countries, studies have shown that even when they are performing at the same levels as their male peers, girls are less confident in their math and science abilities (Correll 2001; Else-Quest et al. 2010). Correll’s (2001) theory of “biased self-assessment” highlights the role of self-assessments

as determinants of career-related decisions. In her longitudinal analysis of U.S. data, Correll found that girls undervalued their own competence in math, which discouraged them from pursuing quantitative coursework and fields of study. This line of research shows more generally that self-assessments in academic domains mediate the influence of gender on high school course choices and career aspirations (Marsh and Yeung 1997; Nagy et al. 2008; Sikora and Pokropek 2012).

We extend research on gender stereotypes by examining the consequences of “biased” self-assessments for career aspirations in the cross-national context. Because of the considerable variation across countries in the size of gender differences in STEM attitudes and behaviors (Baker and Jones 1993; Else-Quest et al. 2010), cross-national comparisons have inherent interest. They also allow us to elaborate and test hypotheses that stem from expectation states theory (Ridgeway, 2011). Expectation states theory explains how cultural status beliefs allow different performance expectations for boys and girls to emerge. Status beliefs lead boys to evaluate their math and science abilities more highly than girls do, even in highly affluent societies where girls have relatively unimpeded opportunities to higher education and to professional labor markets. This happens either because girls believe they are less capable in math and science or because they are aware that “this belief exists in the culture and expect that others will treat them according to it” (Correll 2001:1696). This response is not illogical because teachers can have gender-stereotypical views about math and science that influence the assignments and the encouragement given to students (Riegle-Crumb and Humphries 2012), and stereotypes affect the demand for women in mathematics-related tasks, regardless of ability (Reuben et al., 2014). Because cultural beliefs about gender and STEM suitability are shaped by the environment, we would expect to find variation in gender differences in career aspirations across countries, just as we would expect to find variation across schools and communities (Legewie and DiPrete 2014; Mann et al. 2015).

In this article, we use a Bayesian decision model to formulate testable hypotheses that follow directly from expectation states theory (Foschi and Foschi 1976). This decision model allows us to differentiate ability signals (operationalized as test scores) and preferences (a latent construct). Using data from the Program for International Student Assessment (PISA) 2006, we analyze the importance of gender-stereotyping for gender differences in the formation of STEM aspirations among 15-year-old students. The first step of our analysis measures the variation in science self-assessments across countries and by gender, net of other factors. In the second step, we examine the consequences of the country (residual) self-assessment effect and the gender gap in the self-assessment effect (both constructed from step 1), and the national math-science performance level for gender differences in the development of a STEM orientation. These are our primary measures of the country performance and self-assessment contexts within which boys and girls form STEM aspirations. Then, to establish the unique importance of the country performance and self-assessment environments, we bring in additional measures of school context to differentiate the between-country and within-country effects on gender differences in STEM aspirations.

To anticipate our results, we reconfirm that girls generally are less likely than boys to aspire to STEM occupations, even when they have comparable abilities and

self-evaluations in math and science. We then show empirically that in countries with high average performance and self-assessment, female students demand especially strong performance of themselves in math and science in order to form a STEM orientation.

Our results are of theoretical importance because they show that the decision-theoretic model underlying expectation states theory operates not only at the level of individuals but also at the level of environments. As a consequence (and as we empirically demonstrate), the gendered component of the formation of STEM orientations varies systematically across national environments. We show that—as predicted from theory—the subpopulation of students who express a STEM orientation in secondary school is even more disproportionately male in countries with strong math and science performance than is true in the average country. However, an ironic consequence of the greater sensitivity of girls than boys to the performance environment is that girls with STEM orientations are even more disproportionately drawn from the highest-performing female students in countries with stronger performance in math and science than in countries with weaker performance levels.

Finally, we bring in a traditional measure of country gender equity to identify the distinct effects of the gender culture and the performance environment on gender differences in STEM aspirations. Whereas other studies (Else-Quest et al. 2010; Charles and Bradley 2009; Charles et al. 2014) have emphasized the paradoxical finding that gender differences in STEM are larger in more gender-egalitarian countries and have proposed explanations for this paradoxical finding, we show that the paradox disappears when the performance environment is included in the model. Net of the performance environment, the gender gap in STEM orientations is smaller in more gender-egalitarian countries.

## Country-Level Gender Differences in STEM Performance and Self-Assessment

The standard explanation for gender differences in math and science self-assessments is an environmental, essentialist theory (Charles 2011). Students think that math and science ability is to some extent a fixed trait, that it varies in the population, and that how good one is in math and science can only be judged relative to the performance of others in the environment. The environment includes the other gender, and gender essentialism says that girls are in general inferior to boys in performance in part because they are girls, which leads to generally higher self-assessment of math and science abilities for boys than for girls even when the actual performance of students is taken into account. Boys generally do slightly outperform girls in math and science testing (Hyde and Mertz 2009; Penner 2008), though the reason for this may be nothing other than the performance benefits to boys from their typically higher self-assessments (Cohen and Sherman 2014; Lovaglia et al. 1998; Steele 1997).

Viewing gender as “an institutionalized system of social practices for constituting males and females as different” (Ridgeway 2001:637), sociologists and social

psychologists have demonstrated that gender differences consist in considerable measure of shared stereotypes about femininity and masculinity (Ridgeway and Correll 2004). Expectation states theory argues that those stereotypes contain status beliefs that associate greater competence with boys than girls, particularly in valued social arenas (Correll 2004). As diffuse status characteristics, gender differences influence behavior when they become salient in a particular setting (Correll 2004; Correll and Ridgeway 2006). The expectation states literature, for the most part, examines the way that different performance expectations for group members emerge in collective settings (Berger et al. 1972; Correll and Ridgeway 2006). This theory has been extended to examine the impact of gender status beliefs on self-evaluations (Correll 2001, 2004; Lovaglia et al. 1998), but there has been little theoretical development of the way in which stereotypes interact with observed performance metrics to influence self-assessments and expectations of competence.<sup>1</sup>

In a heavily cited paper, Correll (2001) (see also Correll 2004) found that girls undervalued the implications of their objective performance in mathematics relative to boys and showed that the gender gap in self-assessment produced gender differences in persistence on a path toward a career in science, math, or engineering. Correll's finding of a gender-biased self-assessment agrees with the predicted behavior pattern from expectation states theory and the closely related status characteristics and double standards theories (Correll 2004; Foschi 2000; Ridgeway and Correll 2004; Ridgeway 2011).<sup>2</sup> Girls will be perceived by others and will perceive themselves to be less competent than boys who have the same level of objective performance, and the effect is particularly large for tasks that are male-typed, such as math and science. Correll (2001) further found that the effect (slope) of math performance on self-assessment was larger for girls than it was for boys. As Correll writes, "performance feedback, in the form of math grades, should have a larger impact on females' self-assessments of their mathematical ability, since they must contend with lower societal expectations of their mathematical competency" (2001:1715). The implication is that boys are more likely to think they can perform well in math and science even if their performance is mediocre, because—as boys—they can continue to believe that their math-science ability is high. Correll (2001) did not consider whether there were gender differences in returns to performance or self-assessments for STEM orientations or behaviors. However, her findings are important because they establish a behavioral link between self-assessments in the early secondary school years and decisions in later years.

To understand the role that macro-level belief systems play in the formation of self-assessments and career orientations, it is necessary to consider their operation in multiple contexts. This is quite different from Correll's work, which was conducted in a single country (the United States) at a single point in time.<sup>3</sup> Researchers have found considerable cross-national variation in gender differences in performance levels, which supports the idea that gender gaps in achievement are products of cultural forces (Hyde and Mertz 2009). Scholars modeling that variation have shown that the magnitude of gender gaps varies with specific structural and cultural characteristics of countries. Because cross-national comparisons are riddled with difficult conceptual and measurement issues, it is not surprising that the research has led to inconsistent results (Guiso et al. 2008; Penner 2008; Fryer and Levitt

2010; Kane and Mertz 2012; Hamamura 2012; Nosek et al. 2009). To the extent those studies can be reconciled, they suggest that a country's levels of development and gender equity are associated with smaller gender gaps in math and science achievement, but that gender stereotyping and low power distances produce effects that potentially offset the effects of gender egalitarianism and development. In contrast to these results, scholars modeling cross-national variation in math self-assessments have found that the male advantage in self-assessments is larger in more highly developed and gender-egalitarian countries (Charles et al. 2014; Else-Quest et al. 2010).

Expectation states theory is readily extendable to the case of multiple contexts. Expectation states theory is at its heart a Bayesian perspective on decision-making. Foschi and Foschi (1976) developed a Bayesian model for the formation of performance expectations in which evaluations of the performances by self and other are the data to be processed and standards are the parameter that regulates the formation of expectations. Their model implies that standards for ability will be consistent with status differences such that standards will be stricter for women than for men (Foschi 2000). To elaborate, we suggest that men and women bring prior beliefs—including competence beliefs associated with diffuse status characteristics like gender—to a decision problem. They collect data about the problem and make a decision about behavior based on the strength of the data and the strength of the prior. Gendered beliefs are not a constant either across societies or over time (Charles and Bradley 2009; Ridgeway 2011). Generally speaking, one would expect that the gender gap in evaluations of competence would be larger in environments in which a particular task was more strongly male-typed and also the more strongly that gender was seen to be a status that differentiated behavior in general. One would also expect that empirical evidence would be taken more seriously by women when the environment was providing a stronger prior belief—in the form of discrimination or stereotyping—about the disadvantage of women for a particular task.

Putting aside for the moment the question of measuring the gender culture in the macro environment, the impact of the environment on a potentially gendered behavior such as the formation of a STEM orientation can be seen more rigorously if we provide a theoretical statement of the process that is consistent with but more formally stated than in Correll (2001). We follow the Bayesian formulation used in recent papers on voting and expert decision-making (Hansen et al. 2014; Iaryczower and Shum 2012). The online supplement contains additional technical detail along with an example using simulated data that shows the stability of the model across varying assumptions. Assume that a student has to make a decision about whether to plan to work in a STEM occupation and, therefore, whether to take the necessary coursework and other actions required to major in a STEM field. She wants to know whether she would be suitable and happy in terms of ability and interest for such work. The "utility" of such a decision depends on an unknown state variable that is different for every student. The student knows her performance in math and science courses, which, following the literature, we call a "signal" about that student's unknown suitability for STEM work. Based on this signal, the student has to decide whether to plan a STEM career or not. There are two errors

that the student can potentially make. One is to embark on a STEM orientation when that student would be more suitable (in terms of interest, performance, or happiness, summarized as “utility”) with a non-STEM orientation. The second error would be to choose a non-STEM orientation when she would be more suitable for a STEM orientation. It can be shown that a Bayesian decision-maker would choose a STEM orientation by comparing two quantities. The first quantity is the ratio of the probability that the student is more suitable for a STEM field, given the student’s “signal,” to the probability that she is more suitable for a non-STEM field, given her signal. The second quantity is the ratio of the loss of utility the student suffers by choosing a STEM field when that student is better suited for non-STEM work to the loss of utility the student suffers by choosing a non-STEM field when that student is better suited for STEM work. If the first ratio is larger than the second ratio, the student would choose a STEM orientation, and otherwise, she would choose a non-STEM orientation. In other words, the decision rule is

*Choose STEM if*

$$\frac{\text{Prob}(\omega_i = \text{STEM}|\text{signal}_i)}{\text{Prob}(\omega_i = \text{non-STEM}|\text{signal}_i)} > \frac{\text{cost of mistakenly choosing STEM}}{\text{cost of mistakenly avoiding STEM}} = \frac{\pi_i}{1 - \pi_i} \quad (1)$$

Because performance in high school math and science is an important predictor of a student’s ability to do higher-level STEM work and also of the student’s enjoyment of such work, it follows that the stronger is the math-science “signal,” the higher is the probability that his/her “state” variable is STEM. Because the student evaluates the cost of mistakenly planning a STEM career relative to the cost of mistakenly planning a non-STEM career, it follows that students with stronger STEM signals are more likely to choose a STEM orientation than students with weaker STEM signals. Two additional premises follow from expectation states theory. First, girls are less likely to choose a STEM field even if their loss ratio is the same as that for boys because they too-conservatively read the “signal” they get from their performance, whether because their “prior” beliefs tell them that girls are generally not STEM oriented or because they view their performance signal to be unreliable information about their true orientation. Second, a girl’s loss ratio will generally be higher than that for a boy with the same level of performance. This means that even if the girl reads her “signal” from performance the same way a boy would who had the same performance level, she is less likely than the boy to choose STEM, because she can well imagine that entering a field that is commonly thought to be suited for boys and finding that she is not suitable is a worse mistake than planning for a non-STEM career even if she might be better suited to STEM work. In terms of the formalism of the decision rule, the value of  $\pi$  varies within the population of girls and also within the population of boys, but the average value of  $\pi$  is higher for girls than for boys.

Importantly, however, these fractions and the decisions they produce will depend on environmental characteristics, including the performance of other students in the environment. First, we consider the case of boys as students in an environment in which students in general perform well in STEM courses. In such

environments, we conjecture, boys will perceive STEM fields of study to be highly competitive and challenging because students in general do well in math and science. The frames of reference that individuals use to evaluate themselves will reflect the higher performance level (Davis 1966; Marsh and Hau 2003; Nagengast and Marsh 2012). In a high-performance environment, the risk of failure grows higher for any given level of individual performance because the competition is perceived to be strong. In such environments,  $\pi$  is likely to be relatively high. Boys are likely to require stronger evidence that they are good in math and science before deciding to pursue a STEM orientation. Operationalizing the relationship between true state and signal as a latent index function, we conclude that the intercept becomes smaller in environments with higher math and science achievement and the slope on own performance becomes larger. The consequence of a higher slope is that we expect boys with STEM orientations to be disproportionately drawn from the top of the math and science distribution in countries that have higher-performing students in math and science.

The theory makes clear predictions about how girls and boys respond differently to high STEM-performance environments. In environments in which STEM performance is high, girls, like boys, require stronger evidence that they are good in math and science before deciding to pursue a STEM orientation. But the burden is larger for girls than boys because of the presence of gender-stereotyped ideas about math and science, math and science as fields of study, and math and science careers. Unless the evidence for personal suitability to STEM is very strong, a girl will generally perceive the expected loss from mistakenly choosing a STEM orientation for herself (i.e., guessing that she “belongs” in a STEM field only later to discover that this is not so) will be higher than for a boy, precisely because math and science are male-typed interests, fields, and careers. Personal suitability, it must be emphasized, is a relative judgment; it depends on own performance and also the strength of the performance environment. Consequently, girls demand relatively more evidence than boys in stronger performance environments in which the risk of making a mistake about one’s suitability for math and science (more costly for a girl than a boy) is higher. In the latent index model, this implies that the intercept drops and the slope on own performance rises with the strength of the performance environment more strongly for girls than for boys. The aggregate consequence of this gendered pattern in the impact of performance environments on the coefficients of the latent index models is both that a smaller share of students with STEM orientations are female, and that those girls who do adopt a STEM orientation are strongly selected from the top of the female math and science distribution.

This leads back to the question of how to measure cultural stereotypes in the environment. One obvious aspect of the national gender culture is the extent of gender stratification in the educational and occupational sectors. The gender stratification hypothesis predicts that gender gaps in achievement, self-assessments, and career aspirations will be smaller in egalitarian societies in which greater opportunities exist for women in education and in labor markets (Baker and Jones 1993; Eccles 1994). Because the gender gap in self-assessments is generally larger in more egalitarian countries (Charles et al. 2014; Else-Quest et al. 2010), gender stratification and discrimination are insufficient to explain gender gaps in STEM aspirations and

attitudes. The conventional measures of gender equity tend to be high in advanced industrial societies, which are countries where math and science performance is relatively high and self-assessments are relatively low. Other studies have looked directly at the level of economic development and have found that gender gaps in self-assessment as well as occupational segregation are relatively high in advanced industrial societies (Charles and Bradley 2009; Else-Quest et al. 2010). Collectively, those studies raise the question whether gender stereotypes about math and science strengthen in environments in which math and science achievement is very high and gender stratification is limited. This apparent discrepancy motivates our measurement strategy.

Psychological research provides several related empirically established findings that help to explain the linkage of cultural stereotypes and the performance environment. One such finding is stereotype threat, which occurs when stereotypes directly interfere with performance and expectations of performance. The activation of stereotypes constitutes a threat to self-integrity that hampers performance by arousing stress, causing one to consume resources that otherwise could be used for problem solving; stereotypes also interfere with performance when they transfer one's focus to the short term goal of self-defense (Cohen and Sherman 2014; Steele 1997). There are several reasons to think that gender stereotypes related to math and science are more prevalent in societies in which performance levels are high. Social psychologists have found that stereotype threat is more likely to disrupt girls' math performance when test material is difficult (Spencer et al. 1999; Steele 1997) or when test takers care a great deal about their math abilities (Aronson et al. 1999). Because the national performance environment reflects both the difficulty of the curriculum and the importance of math and science fields in a given country, we suggest that the aggregated performance level is a proxy for the likelihood that gender stereotyping occurs.

Another large body of research, including experimental studies, has found that women often respond less favorably to competition than men (Bönte 2015; Gneezy et al. 2003; Niederle and Vesterlund 2007, 2010).<sup>4</sup> Buser et al. (2014) found that a sizable part of the gender gap in the choice of the most math heavy curricular track in the Netherlands (after accounting for grades) was explained by the relatively greater levels of competitiveness among boys. Similarly, Alon and DiPrete (2014) found that the intensity of competition as measured by public information about the (changing) standards for admission to specific majors has a larger deterring effect on female than male applicants to STEM fields at elite Israeli universities. The role of competition goes beyond simple risk aversion on the part of women. Experimental research suggests that men are more likely than women to overreach in aiming for more challenging tasks that go beyond their abilities (Buser et al. 2014; Niederle and Vesterlund 2010) and that competition has a stronger deterring effect on relatively lower-performing women than on lower-performing men (Niederle and Vesterlund 2007). Those studies collectively suggest that the pure intensity of competition in high-performing environments might stimulate gender gaps in self-assessment apart from or in addition to the strength of cultural beliefs about gender's link to particular performance tasks.

Sociological research provides another explanation for larger gender gaps in STEM attitudes in environments in which math and science achievement is very high and gender stratification is limited—one that does not run directly through gender stereotypes about math and science, but through ideals of self-expression (Cech 2013). Recent studies observe that the more affluent, gender-egalitarian societies have cultures that “celebrate gender difference and simultaneously endorse self-expression” and structural features of their higher education systems and occupational sectors that support gendered curricular affinities (Charles and Bradley 2009:927). Charles and colleagues suggest that the difficulty and technical nature of the math and science curriculum in high-performing country environments has a disproportionate tendency to weaken girls’ interest in those fields of study and to strengthen their interest in “artistic or human-centred pursuits” (Charles et al. 2014:88). Although this mechanism may operate in some circumstances, we show below that this surprising pattern of larger attitudinal gender gaps in more egalitarian societies reverses when we account for the aggregate performance and self-assessment levels in the country.

Nevertheless, consistent with other researchers (Charles and Bradley 2009; Cotter et al. 2011; Cech 2013), we argue that egalitarian cultures do not eliminate gender stereotyping, even as they alter its content. Even in a relatively gender-egalitarian culture, expectation states theory remains valid. We expect young boys and girls to measure their abilities against both the local and the global environment. Generally, strong performance by other students in a school environment reduces the assessment of one’s own abilities for both boys and girls (Marsh and Hau 2003; Nagengast and Marsh 2012). In such environments, both boys and girls require more evidence that they themselves are strong performers to overcome the stronger prior assumption that they are not talented. This evidence requirement produces a stronger self-assessment return to performance (i.e., a higher slope of performance in the self-assessment model) and a stronger STEM orientation return to both measured performance and self-assessment. Following expectation states theory, we expect those effects to be especially strong for girls. Stronger country performance implies more difficult curricula, higher performance standards, and greater competition, all of which have been shown to heighten gender-essentialist ideas about math and science. In such environments, girls must confront the initial stereotype that they are not as good at math as boys and the consequent higher perceived cost to them of making the mistake of doing STEM when it is not the “right” choice for them. Ultimately, we need not resolve the precise mechanisms by which performance levels enhance gender stereotypes. There are many possible mechanisms for this pattern and, in the absence of systematic measures of gender culture at the country level, we do not place exclusive weight on any particular factor. Our theory and empirical results are robust to the precise mechanisms by which the strength of the performance environment produces gender-specific reactions when the performance tasks in question are gendered.

In the analysis that follows, we test this theory by using the national performance level in math and science as well as the country-level self-assessment effect (after controls for individual performance and individual family background) as measures of the strength of math and science performance and self-assessment environments.

We follow Charles and Bradley (2009:938) in using the aggregated country-level gender gap in the self-assessment effect “as an indicator of the extent to which mathematics and mathematics curricula are culturally gendered.” We justify the use of these aggregate measures by comparing them to other indicators of gender cultural beliefs. We then demonstrate that the pattern of response by male and female students to the environment follows our multilevel generalization of the predictions of expectation states theory to the case of varying national environments.

## Data and Methods

We use data from the Program for International Student Assessment (PISA), which is a triennial international study that tests the reading, mathematical, and scientific literacy level of 15-year-old students who are still in school. The database is hierarchically structured such that students are nested within schools, and schools are nested within countries. We use the 2006 data collection, which included 57 countries. In 2006, science was the major content domain. We exclude data from Qatar because the students were not asked about STEM aspirations. Although the PISA 2012 data recollection is more recent, we do not use it for the main analyses because it does not include an open-ended measure of occupational aspirations.<sup>5</sup>

### *Self-Assessment*

The student questionnaire measured attitudes and motivational orientations towards science. Attitude scales were developed from 45 questions grouped into seven constructs—self-concept, self-efficacy, interest, enjoyment, instrumental motivation, future-oriented motivation, and general value. We use the self-concept scale as the dependent variable in the first part of the analysis because it is the measure of self-assessment that is most closely associated with motivation (Ferla et al. 2009). The self-concept scale consists of six equally weighted items, each of which used a 4-point scale of agree–disagree responses: “Learning advanced science topics would be easy for me,” “I can usually give good answers to test questions on science topics,” “I learn science topics quickly,” “Science topics are easy for me,” “When I am being taught science, I can understand the concepts very well,” and “I can easily understand new ideas in science.”

### *STEM Aspirations*

The dependent variable in the second part of the analysis is whether the student expects to have a STEM job at the age of 30. The question taken from the student questionnaire was “What kind of job do you expect to have when you are about 30 years old? Write the job title—.” The responses were coded using the International Standard Classification of Occupations; the detailed list of occupations included in the STEM category is presented in the online supplement. Our definition excludes some of the occupations that have been treated as STEM occupations in previous research—namely nursing and occupations at the level of associate or technician and below—because we are interested in a measure of aspirations for elite STEM

careers among high-performing students.<sup>6</sup> Some countries are missing a significant proportion of the occupation data; the proportion of missing data on this variable exceeds 30 percent in Azerbaijan, Israel, and Kyrgyzstan.<sup>7</sup>

### *Performance*

We use the composite scores for math, science, and reading as the basis for the performance variables that we construct. Because of an error in the test booklet, reading test scores from the United States were disqualified. So that we could both use reading test scores as an independent variable in the two steps of the analysis below and keep the United States in those analyses, we substituted data for the United States from PISA 2003, in which mathematics was the major content domain. Math and science self-concepts are highly correlated at the country level. Nevertheless, to test the effects of this substitution, we created separate datasets for the 40 countries that were in both PISA 2003 and 2006, in which we substitute into the pooled models the 2003 data for one country at a time. We then compared the coefficients across those models. The results suggest that the findings in this article are robust to this substitution.<sup>8</sup>

We average the composite math and science scores for each student.<sup>9</sup> We then standardize the average math-science composite score by the mean of the clustering unit (zero sample mean and unit sample variance) such that respondent test scores are centered by school mean test scores, school mean test scores are centered by country mean test scores, and country mean test scores are centered by the grand mean. By standardizing the test score variables by the mean of the clustering unit, we separate the contextual effects into levels, ensuring that the resulting coefficients reflect within-group effects. This also stabilizes the models by removing correlations. We use interaction terms to account for the relation of contextual performance variables to individual performance. We created similar measures using the composite reading scores, which we standardized by the mean of the school test score.

The PISA does not contain information about grades, so test scores are the best measure of performance. The absence of grades is problematic in the sense that students in lower-ability schools tend to have higher grades than students of the same ability levels in higher-ability schools as a result of grading on a “curve” (Marsh 1987). This means that the inverse relationship between ability and self-concept is mediated by the effect of school ability levels on grades. The absence of grades from the regression model predicting self-concept, however, potentially understates the significance of gender differences in self-concepts because girls typically earn higher grades than boys net of ability (Voyer and Voyer 2014).

### *Individual Covariates*

We also use demographic information about each respondent: sex, immigrant status, a broad measure of socioeconomic status (ESCS), and an indicator for whether either parent has a science-related career.<sup>10</sup> Moreover, because countries vary in the extent to which students follow divergent tracks for academic and vocational studies, we include measures for whether the curricular content of the student’s program is

vocational or academic.<sup>11</sup> Because PISA is an age-based sample, we include the student's age at the time of the test (month and year), centered. We also include the student's grade level relative to the modal grade for the country in which the student lives; it is not uncommon in this dataset for students within a country to be enrolled in more than one grade.

### *Country Covariates*

Because of the difficult conceptual and measurement issues involved in making cross-national comparisons, we use multiple indicators of gender culture and development aside from the school and country performance measures that we aggregated from individual test scores. We use the net country effect on self-assessment and the net country gender gap in self-assessment, both of which were estimated (as explained in more detail below) by regressing science self-assessment on a set of measured covariates at the individual level along with country fixed effects. We also use measures of the structural features of the nation's education system as they pertain to the tracking of students between schools; in particular, we use a binary measure for whether assignment into tracks occurs before the age of 16.

To assess the independent contributions of school and national performance levels on self-assessment and STEM aspirations, we also include in our analyses several measures of gender culture that have been used in the literature. We use the Gender Gap Index (GGI), published by the World Economic Forum, and the Gender Empowerment Measure (GEM), published by the United Nations Development Program, as alternative measures of country-level gender equality. The Gender Gap Index assesses gender equity in economic, educational, political, and health domains, and the GEM measures gender equality based on women's political participation and decision-making power, economic participation, and command over resources. In both measures, high values represent greater levels of equality. We compare those measures and the aggregate measures that we create to a series of other measures that are publicly available for select countries. These other country measures include gender equity indices, GDP per capita, Inglehart's self-expression score (Inglehart and Welzel 2005), Hofstede's power distance index (Hofstede et al. 2010), the average gender-math stereotype measured through implicit association tests (Miller et al. 2014; Nosek et al. 2009), and the proportion of women in science fields of study and occupations. For each measure, we used data collected in 2006 or as close as possible to 2006. The online supplement contains a correlation matrix that shows the bivariate relationships among these measures; it also displays the sample sizes for each measure.

### *Weighting*

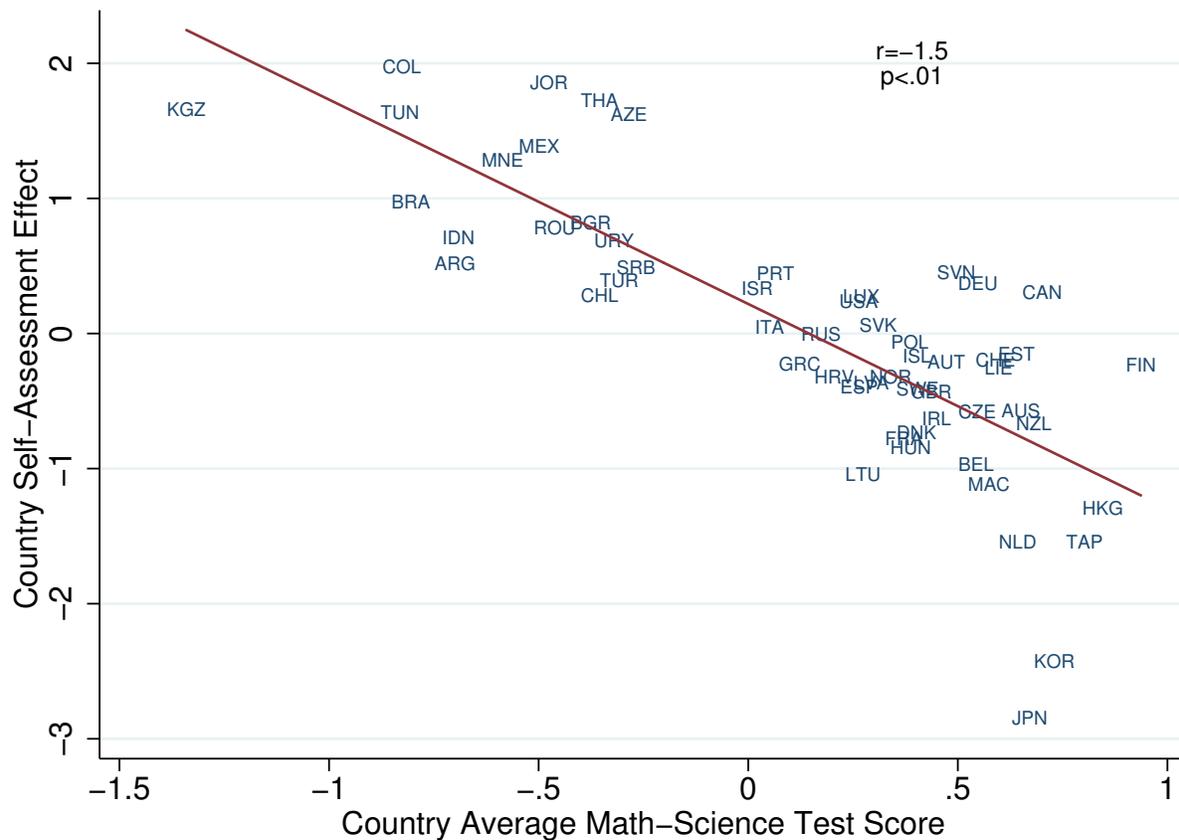
Because PISA has a complex, two-stage stratified sample design, all statistics are weighted using the student-level weights and 80 BRR replicate weights provided in the dataset to compensate for unequal selection probabilities of students. Country sample sizes, weighted sample sizes, and descriptive statistics for the key variables are reported in the online supplement.

## Results

Our first step is to estimate separate models for male and female students in which self-assessment in science is the dependent variable and the independent variables are performance in math and science, performance in reading, individual covariates, school and country performance variables and their interactions with individual performance, and a fixed effect for the country. The country fixed effect captures all aspects of the country's environment that produce a gap in self-assessment between boys and girls that cannot be explained by the individual or school characteristics included in the specification. By differencing the country effects for boys and girls, we obtained a country measure for the gender gap in self-assessment that is independent of performance (CntGG). We then combined the male and female samples and estimated a gender-combined country effect on self-assessment (CntSAEff).<sup>12</sup> We call these measures the net country effect on self-assessment and the net country gender gap in self-assessment. They are country scores estimated with fixed effects that are net of a set of measured covariates at the individual level. We standardized the measures (mean=0, SD=1), such that higher values for the CntGG represent a larger male advantage. We use these country self-assessment scores, along with country performance levels, to test the hypotheses that we discussed above.

We argued that countries with high levels of math and science achievement would have relatively low levels of science self-assessment (net of performance).<sup>13</sup> To test this argument, we created a scatterplot showing the relationship between the country average math-science test score and the gender-combined country effect on self-assessment. This relationship, which is displayed in Figure 1, is clearly negative and statistically significant; countries with higher average math and science scores have lower average self-assessment scores. Following expectation states theory, we hypothesize that high achievement/low self-assessment environments should be associated with relatively large gender gaps in self-assessment. To test this hypothesis, we plot the relationship between the size of the country self-assessment effect and the size of the gender gap in self-assessment in Figure 2. These results show that, as predicted, the gender gap in self assessment is larger in countries with more negative self-assessment effects. Figure 3 further demonstrates that country self-assessment effects for both girls and boys vary inversely with country performance, but that the relationship is stronger for women. As a consequence, the gender gap in self-assessment is relatively large in countries that have high average levels of math and science performance. All three of these associations match the theoretical prediction.

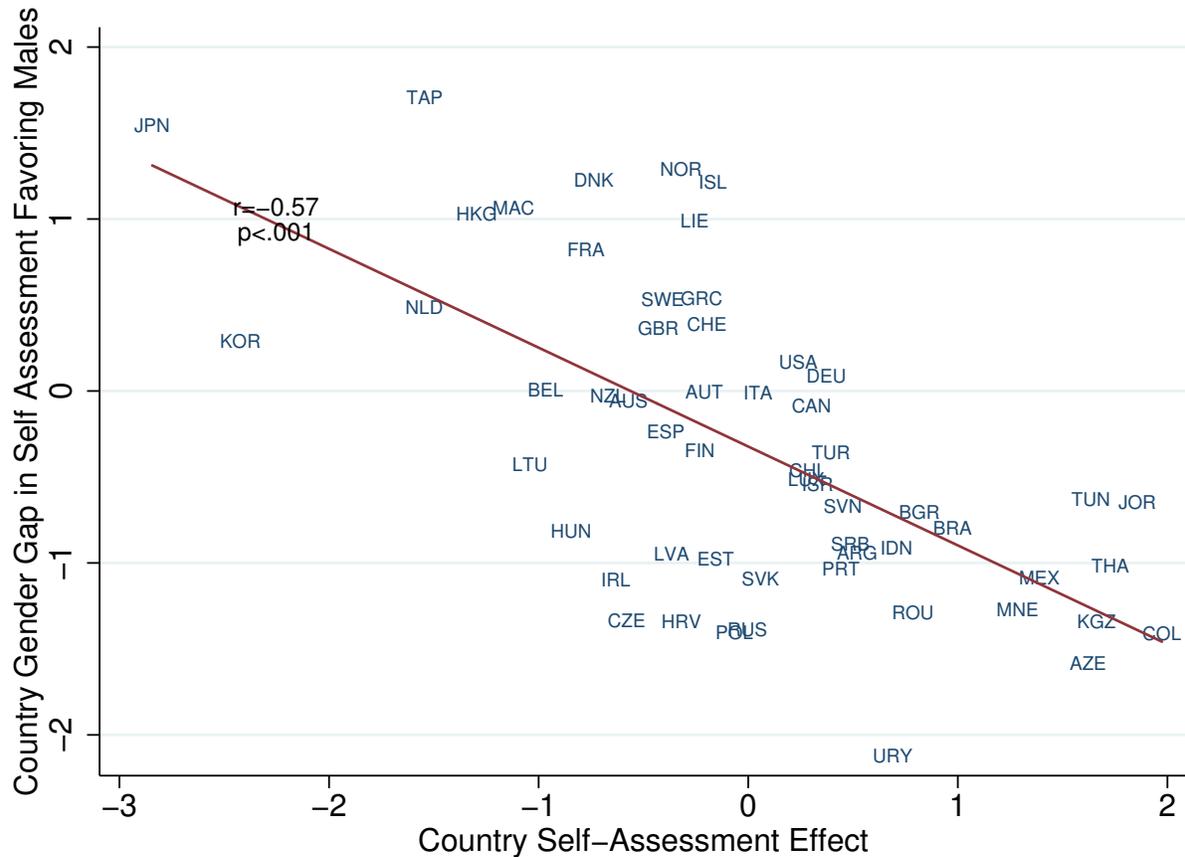
To summarize, Figures 1, 2, and 3 collectively provide strong support for our first hypothesis; both boys and girls experience more negative self-assessment in countries where the average achievement levels are high, but the effect is stronger for girls. The mechanism for this effect, we conjecture, is that high average performance in math and science in a country enhances the stereotype that high math and science ability is a male trait. This conjecture is supported by the relationships among these three variables and other measures of gender equity. As shown in the online supplement, commonly used measures of overall gender equity—the GEM



**Figure 1:** Country self-assessment effect versus country average math and science achievement.

and the GGI—are positively associated with country performance and the gender gap in the country residual self-assessment effect. Measures of STEM parity—the percent of female science graduates, the percent of women in tertiary education that is enrolled in science fields of study, and the percent of females engaged in science occupations—are negatively correlated with country performance and the gender gap in the country residual self-assessment effect.

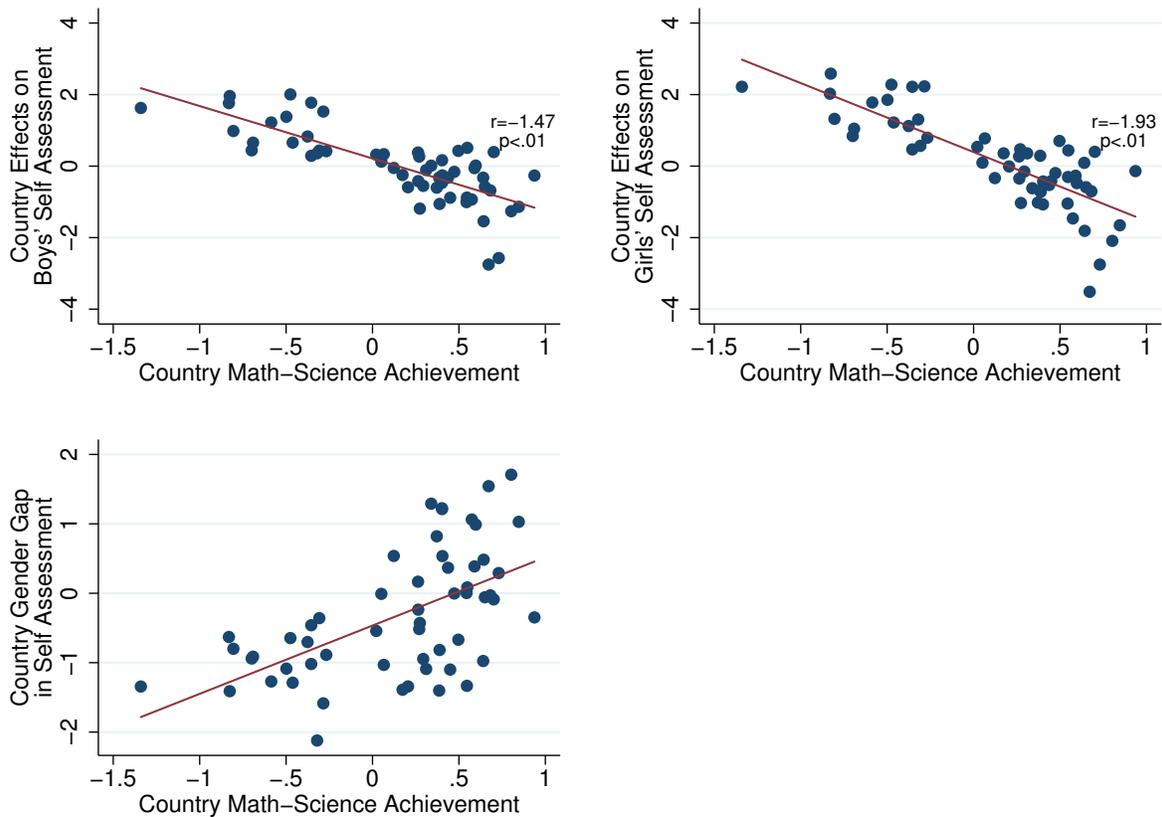
To further assess the validity of our expectations, we estimate random-effects models for boys and girls (comparable to the fixed-effects models discussed above), with the addition of the main effect of country-level math-science test scores and the country Gender Gap Index for 2006. Table 1 contains the results. As predicted, both boys and girls have lower self-assessments in countries where the average level of math and science achievement is high. Boys and girls have slightly higher self-assessments in gender-egalitarian countries, net of individual and contextual performance measures, but girls' self-assessments increase to a lesser extent than those of their male peers (the paradoxical widening of the gender gap in self-assessments reported by Else-Quest et al. [2010]). The results are substantively the same when we use the 2006 GEM as a measure of gender egalitarianism.



**Figure 2:** Gender gap in self-assessment versus country self-assessment effect.

## Environmental Effects on the Slope of Achievement and Self-Assessment

Research in the United States has demonstrated the negative consequences of low self-assessment for the likelihood of forming a positive orientation and a behavioral response towards STEM fields of study and occupations (Correll 2001). This result is readily explainable in terms of expectation states theory. We next examine whether the same finding is generalizable to the cross-national context. One would certainly expect that an individual's low self-assessment would depress his/her probability of developing a positive STEM orientation, and therefore one would expect that a country with a larger self-assessment effect would have a higher proportion of students aspiring to STEM careers. One also would expect that in low self-assessment, high national performance environments, students would hold themselves to stricter standards before forming STEM orientations, so the returns to math-science test scores would be higher. Students in these countries require more evidence of their own capability in math and science before they are willing



**Figure 3:** Country self-assessment effect versus country math and science performance.

to develop a STEM orientation because they believe that one needs to be especially good at math and science in order to enter STEM fields and STEM occupations.

Our next set of models tests the hypothesis that high-achievement environments raise the STEM orientation returns to math achievement (even though—as shown above—they depress overall levels of self-assessment). We use multilevel models with random intercepts for schools and countries to take into account the clustered structure of the data. We fit a model in which orientation towards a future STEM occupation was the dependent variable. We modeled STEM orientation as a function of own achievement, the level of achievement in the school and national context, and an interaction effect between the country average achievement and own achievement. We then conducted an additional analysis in which we substituted for the above interaction effect an interaction effect between the country self-assessment effect and own achievement. The results of these models are in Table 2.

As we predicted, individual-level STEM orientation increases with higher individual performance in math-science and reading, but the country achievement level decreases individual-level STEM orientations. Students in countries where

**Table 1:** Regression of Self-Assessment on Individual- and Country-Level Measures and Random Country Effects

	(1) Female	(2) Male
Reading Score (R)	-0.098 <sup>†</sup> (0.005)	-0.079 <sup>†</sup> (0.005)
Math-science score (MS)	0.265 <sup>†</sup> (0.004)	0.252 <sup>†</sup> (0.004)
School average math-science score (SchMS)	0.104 <sup>†</sup> (0.007)	0.110 <sup>†</sup> (0.007)
Country average math-science score (CntMS)	-0.620 <sup>†</sup> (0.077)	-0.531 <sup>†</sup> (0.064)
MSxCntMS	0.232 <sup>†</sup> (0.005)	0.272 <sup>†</sup> (0.005)
GGI 2006	1.081 (0.825)	1.517* (0.687)
Constant	0.025 (0.037)	0.125 <sup>†</sup> (0.030)
Number of Observations	146,198	131,861

Controls include socioeconomic status, having a parent with a science-related career, being an immigrant, age, relative grade level, and being enrolled in a vocationally oriented program. Standard errors in parentheses.

<sup>†</sup>p < 0.01; \*p < 0.05.

math-science achievement is higher have higher returns to individual math and science performance (model 1). Similarly, the country self-assessment effect increases individual-level STEM orientations, but the returns to individual performance are lower in countries with higher average net levels of self-assessment, reflecting the presumption that students in countries with higher self-assessment require less evidence of their own abilities in order to enter STEM fields (model 2).

## Gender Differences in Response to the Achievement and Self-Assessment Environment

Next, we address the issue of gender differences in response to the achievement and self-assessment environment. We provided evidence in the previous sections that high math-science achievement cultures would lower self-assessment, net of own performance. We also hypothesized that in high-achievement, low self-assessment cultures, girls more than boys would doubt their competence in math and science and would require greater evidence of their competence in order to develop a STEM orientation. To test this last hypothesis, we combined the performance and self-assessment measures and elaborated our logistic regression models to allow for interactions between individual achievement and country math and

**Table 2:** Models for STEM Orientation Including Interaction Effects Between the Environment and Individual Math-Science Performance

	(1) Performance	(2) Self-Assessment
Reading score (R)	−0.029 <sup>†</sup> (0.009)	−0.003 <sup>†</sup> (0.009)
Math-science score (MS)	0.453 <sup>†</sup> (0.007)	0.495 <sup>†</sup> (0.007)
School average math-science score (SchMS)	0.522 <sup>†</sup> (0.014)	0.499 <sup>†</sup> (0.014)
Country average math-science score (CntMS)	−0.792 <sup>†</sup> (0.108)	−0.584 <sup>†</sup> (0.172)
MSxCntMS	0.276 <sup>†</sup> (0.009)	
Country self-assessment effect (CntSAEff)		0.135 (0.091)
MSxCntSAEff		−0.113 <sup>†</sup> (0.005)
Constant	−1.449 <sup>†</sup> (0.061)	−1.479 <sup>†</sup> (0.064)
Number of observations	324,512	

Controls include socioeconomic status, having a parent with a science-related career, being an immigrant, age, relative grade level, and being enrolled in a vocationally oriented program. Standard errors in parentheses.

<sup>†</sup>p < 0.01; \*p < 0.05.

science achievement, and between individual achievement and the country net self-assessment effect. To show the three-way interactions of those variables with gender, we estimate separate models for boys and girls. Table 3 shows the results.

The exponentiated intercepts in Table 3 (0.21 versus 0.25) suggest that girls are less likely than boys to have STEM orientations when the independent variables are set at zero.<sup>14</sup> The gap widens in favor of boys in stronger country math-science environments. However, girls have stronger returns to math-science performance, and in stronger math-science environments, the female advantage in the returns to own performance widens; in other words, the female response to a stronger math-science performance environment is larger than is the male response. Girls receive even greater returns to own achievement in stronger self-assessment environments. Girls thus are more sensitive than boys to changes in their own performance levels in these high-performance and high self-assessment cultures.

**Table 3:** Models for STEM Orientation Including Interaction Effects Between Gender, Individual Achievement, and National Math-Science Performance or Self-Assessment Effect

	(1) Female	(2) Male
Reading score (R)	−0.016 (0.014)	0.152 <sup>†</sup> (0.014)
Math-science score (MS)	0.399 <sup>†</sup> (0.011)	0.349 <sup>†</sup> (0.011)
School average math-science score (SchMS)	0.451 <sup>†</sup> (0.019)	0.396 <sup>†</sup> (0.018)
Country average math-science score (CntMS)	−0.741 <sup>†</sup> (0.164)	−0.466* (0.202)
MSxCntMS	0.339 <sup>†</sup> (0.021)	0.246 <sup>†</sup> (0.022)
Country self-assessment effect (CntSAEff)	0.135 (0.086)	0.104 (0.106)
MSxCntSAEff	0.040 <sup>†</sup> (0.012)	0.014 (0.012)
Constant	−1.518 <sup>†</sup> (0.061)	−1.360 <sup>†</sup> (0.075)
	170,040	154,472

Controls include socioeconomic status, having a parent with a science-related career, being an immigrant, age, relative grade level, and being enrolled in a vocationally oriented program. Standard errors in parentheses.

<sup>†</sup>p < 0.01; \*p < 0.05

## Gender Differences in the Response to the Multilayered State and National Performance Context

Just as STEM aspirations are influenced by gender stereotyping in the national culture, we would expect that gender differences in STEM aspirations would be affected by the performance of peers in the local school environment. To put it differently, the outcomes seen thus far might be expected within a single country when comparing across lower-performing and higher-performing schools. If the same relationships and interactions are observed within-country, when comparing across schools, for example, we might not be identifying the effects of variations in gender-essentialist assumptions embedded in national culture, but rather those “national” effects might be an artifact of school cultures. In this section, we identify separately school peer effects and the national gender culture by considering the gendered response to school and country performance environments. We elaborated our logistic regression models to allow for three-way interactions among the performance variables—test scores at the individual, school, and country levels. We estimate separate models for boys and girls. Table 4 shows the results.

**Table 4:** Models for STEM Orientation Including Interaction Effects Between Gender, Individual Achievement, and National and School Math-Science Performance

	(1) Female	(2) Male
Reading score (R)	−0.023 (0.014)	0.136 <sup>†</sup> (0.014)
Math-science score (MS)	0.409 <sup>†</sup> (0.012)	0.365 <sup>†</sup> (0.011)
School average math-science score (SchMS)	0.450 <sup>†</sup> (0.019)	0.403 <sup>†</sup> (0.019)
Country average math-science score (CntMS)	−0.811 <sup>†</sup> (0.167)	−0.481* (0.205)
SchMSxCntMS	0.388 <sup>†</sup> (0.025)	0.337 <sup>†</sup> (0.025)
MSxSchMS	−0.028* (0.011)	−0.049 <sup>†</sup> (0.011)
MSxCntMS	0.338 <sup>†</sup> (0.021)	0.249 <sup>†</sup> (0.022)
MSxSchMSxCntMS	−0.049* (0.020)	−0.033 (0.020)
Country self-assessment effect (CntSAEff)	0.128 (0.088)	0.113 (0.107)
MSxCntSAEff	0.032 <sup>†</sup> (0.012)	0.007 (0.012)
Constant	−1.523 <sup>†</sup> (0.062)	−1.387 <sup>†</sup> (0.075)
	170,040	154,472

Controls include socioeconomic status, having a parent with a science-related career, being an immigrant, age, relative grade level, and being enrolled in a vocationally oriented program. Standard errors in parentheses.

<sup>†</sup>p < 0.01; \* p < 0.05.

With the fully interacted performance and self-assessment models, the gender differences in the effects of the country math-science test score on the STEM aspirations intercept and individual math-science slopes remain about the same as in the base model. Similarly, gender differences in the response to the self-assessment culture are comparable to those previously reported. Thus, the country effects are not explained by peer influences in the local environment. School effects work somewhat differently than do national environment effects. Unlike at the national level, the female intercept increases (relative to males) in stronger schools, and girls and boys receive about the same returns to own performance in stronger schools.

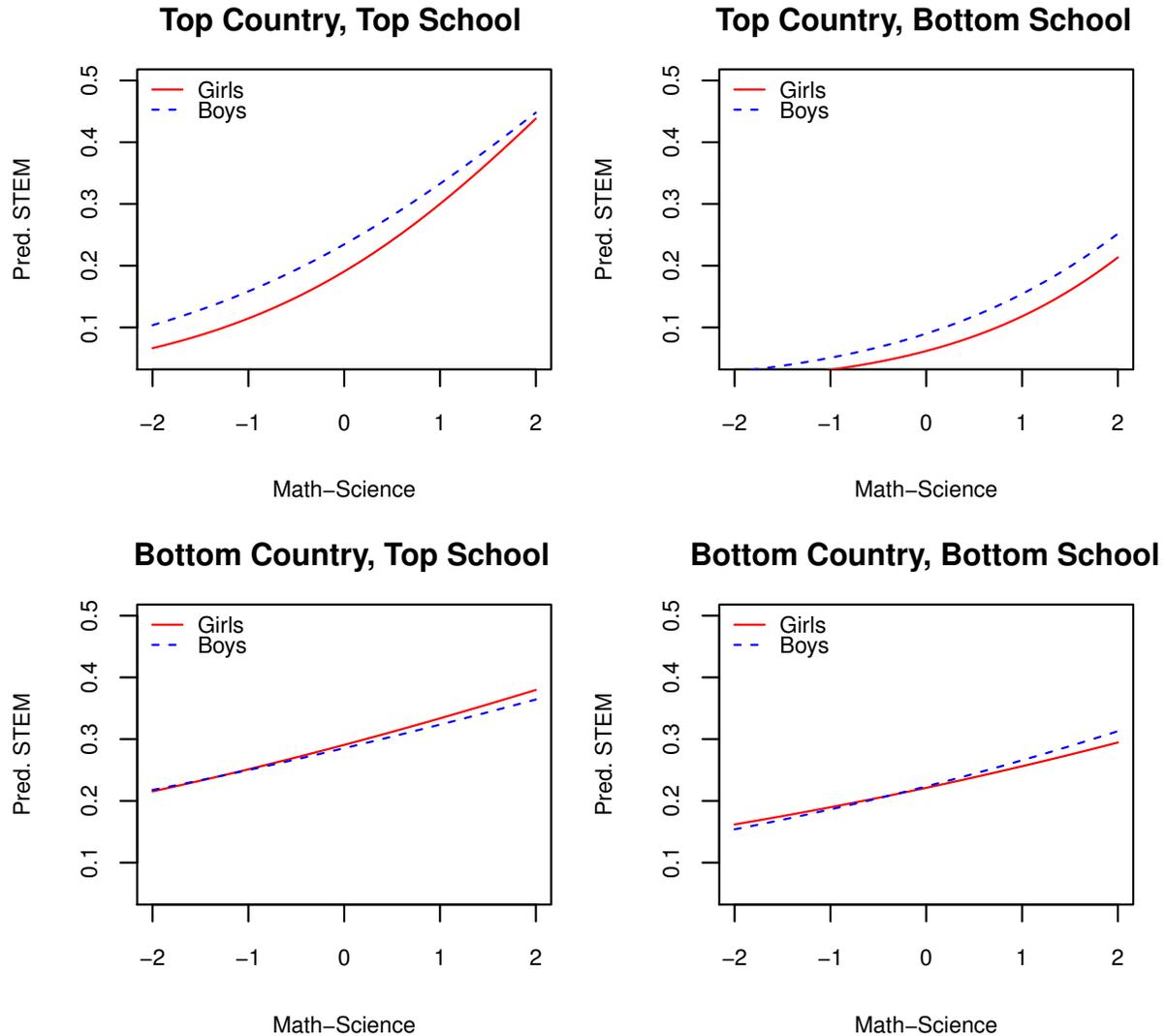
The interactions are presented graphically in Figure 4, which displays the predictive margins on the math-science slope for girls and boys according to the national math-science performance level and the school math-science performance level.

The four panels show how the predicted probability of aspiring to work in a STEM occupation responds to own math-science performance in strong and weak national performance environments and in strong and weak school performance environments. A strong environment represents a country or school at the 90th percentile of the country or school math-science performance distribution, and a weak environment represents a country or school at the 10th percentile of the country or school math-science performance distribution. Each line in these figures shows how the predicted probability responds to own math-science performance, and the different lines in each figure depict this response curve for different genders and in different environments. The comparison of the top and bottom panels depicts the country effects. The predicted probability of a girl forming a STEM orientation is the most responsive to individual math and science performance in strong country and school environments compared to all other environments. Without regard to the type of school environment, the slopes for boys and girls are stronger in strong country performance environments (but girls to a relatively greater extent), and the gender difference in intercepts across national performance environments is easily observed. Overall, the primary effect of strong schools is to raise the overall STEM intercepts for boys and girls alike.

## Using Additional Measures of National Culture to Measure Gender Differences in STEM Aspirations

We next estimated a similar model to that reported in Tables 3 and 4, but this time we included several additional measures of national culture. First, we included a three-way interaction between gender, individual achievement, and the gender gap in the net country self-assessment level (i.e., the difference between the male and the female country self-assessment level, net of individual predictors). The premise is that the gender gap in the net self-assessment country score is an additional measure of the gender stereotyping beyond the country-level math-science achievement and the net country-level self-assessment in science. If this additional measure picks up the strength of gender stereotypes about math and science in the national environment, then we would expect variation in how girls and boys respond to own math-science performance to follow the pattern we have seen above. Second, we include the 2006 Gender Gap Index (GGI) as a measure of gender egalitarianism in the society. The premise is that if gender stratification and gender culture work in different directions, we should be able to separately identify both effects by including the GGI in the model alongside the other national measures. Third, we include a measure of the structural features of the nation's education system as they pertain to the tracking of students before the age of 16. Table 5 displays those results.

Table 5 shows that girls experience relatively larger returns to individual math-science performance in countries with larger gender gaps in net self-assessment. This is in addition to the effects of the country self-assessment level and the national math-science performance level on the returns to own performance (both of which continue to favor girls). This provides further confidence to our general conclusion



**Figure 4:** Predictive margins for boys and girls on the STEM orientation returns to math-science performance by the state and national performance context.

that girls pay relatively more attention to their own performance when forming a STEM orientation in countries where overall performance is high, overall self-assessment is high, and where girls are arguably therefore more sensitized to gender stereotypes about the suitability of math and science careers for women.

Figures 5 and 6 show the predictive margins on the math-science slope for girls and boys according to the country self-assessment effect and the country self-assessment gender gap, respectively. The four panels of Figure 5 show how the predicted probability of aspiring to work in a STEM occupation responds to own math-science performance in strong and weak national performance envi-

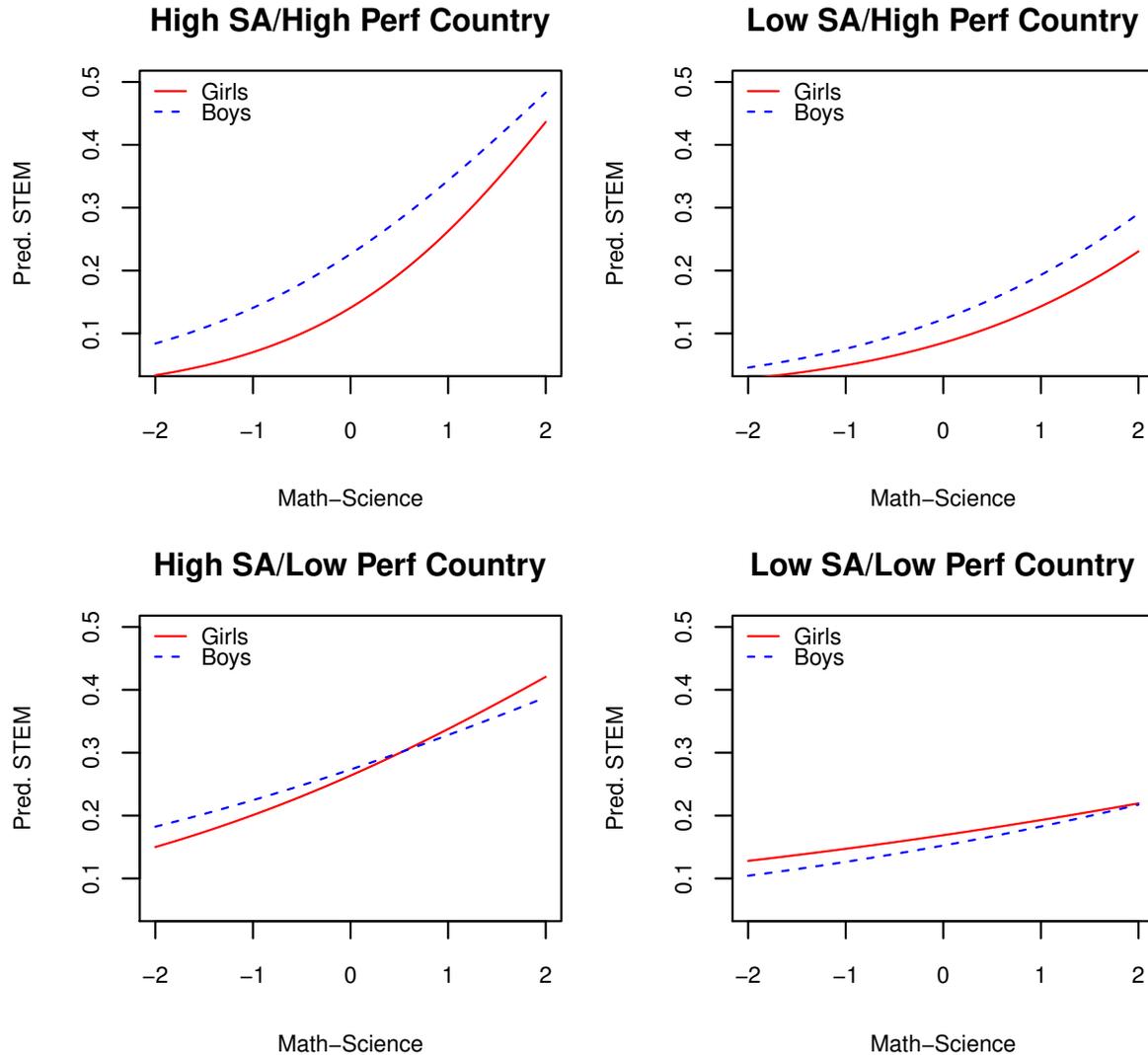
**Table 5:** Models for STEM Orientation Including Interaction Effects with the Gender Gap in the Self-Assessment Effect

	(1) Female	(2) Male
Reading score (R)	−0.008 (0.015)	0.156 <sup>†</sup> (0.015)
Math-science score (MS)	0.442 <sup>†</sup> (0.014)	0.384 <sup>†</sup> (0.014)
School average math-science score (SchMS)	0.441 <sup>†</sup> (0.020)	0.400 <sup>†</sup> (0.020)
Country average math-science score (CntMS)	−0.583 <sup>†</sup> (0.149)	−0.187 (0.200)
SchMSxCntMS	0.386 <sup>†</sup> (0.026)	0.379 <sup>†</sup> (0.027)
MSxSchMS	−0.028* (0.012)	−0.051 <sup>†</sup> (0.011)
MSxCntMS	0.315 <sup>†</sup> (0.022)	0.238 <sup>†</sup> (0.022)
MSxSchMSxCntMS	−0.039 (0.021)	−0.032 (0.021)
Country self-assessment effect (CntSAEff)	0.195* (0.079)	0.279 <sup>†</sup> (0.101)
MSxCntSAEff	0.074 <sup>†</sup> (0.013)	0.019 (0.013)
Country gender gap in self-assessment (CntGG)	−0.029 (0.069)	0.009 (0.087)
MSxCntGG	0.102 <sup>†</sup> (0.013)	0.054 <sup>†</sup> (0.013)
Gender gap index (GGI)	−3.199 <sup>†</sup> (1.125)	−5.433 <sup>†</sup> (1.804)
No country tracking before age 16	0.417 <sup>†</sup> (0.097)	0.274* (0.125)
Constant	−1.741 <sup>†</sup> (0.085)	−1.531 <sup>†</sup> (0.108)
Number of observations	154,238	139,086

Controls include socioeconomic status, having a parent with a science-related career, being an immigrant, age, relative grade level, and being enrolled in a vocationally oriented program. Standard errors in parentheses.

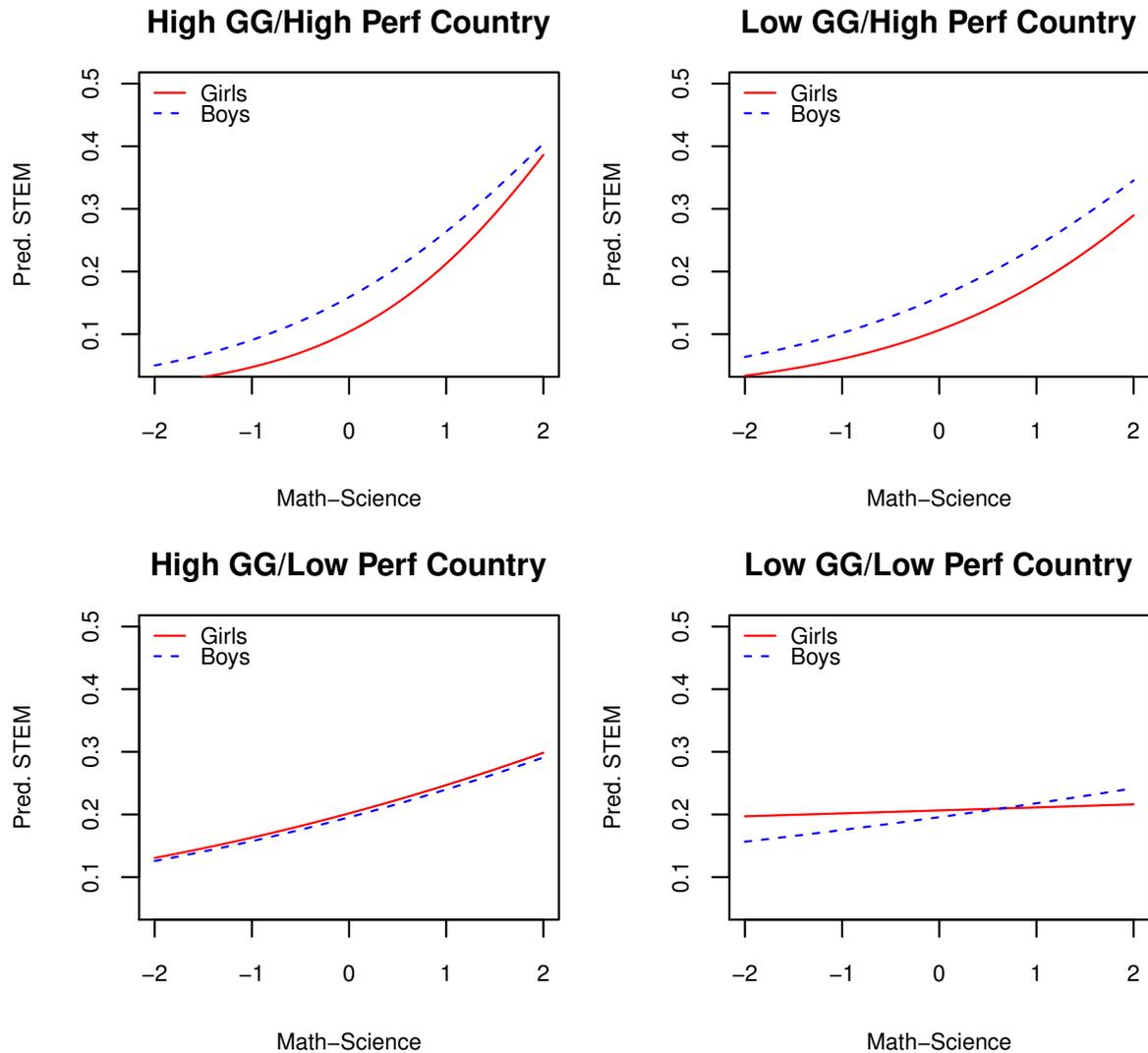
<sup>†</sup>p < 0.01; \*p < 0.05.

ronments and in strong and weak self-assessment environments. Performance environment and self-assessment thresholds are set at the 90th and 10th percentiles. As shown, girls receive relatively greater returns to their own math-science performance in high self-assessment country environments regardless of the performance



**Figure 5:** Predictive margins for boys and girls on the STEM orientation returns to math-science performance by the national self-assessment effect.

environment. The four panels of Figure 6 show how the predicted probability of aspiring to work in a STEM occupation responds to own math-science performance in strong and weak national performance environments and in countries with larger and smaller gender gaps in self-assessment environments. The performance environment and self-assessment gender gap thresholds are set at the 90th and 10th percentiles. As shown, girls receive relatively greater returns to their own math-science performance in countries with larger gender gaps in self-assessment regardless of the performance environment.



**Figure 6:** Predictive margins for boys and girls on the STEM orientation returns to math-science performance by the national gender gap in the self-assessment effect.

Furthermore, the difference in the relative size of the coefficients on GGI ( $-3.2$  for girls versus  $-5.4$  for boys; see Table 5) suggests that the gender gap in STEM aspirations narrows in gender-egalitarian societies after controlling for performance measures. This finding helps to explain the perplexing evidence that gender gaps in STEM aspirations are stronger in more gender-egalitarian countries. The math-science performance level is highly correlated with national levels of gender equality. Once the intensity of the math-science performance culture is taken into account, however, the gender equality measure behaves in a predictable way. In an alternative specification, we substituted Inglehart's self-expression scores (in lieu of

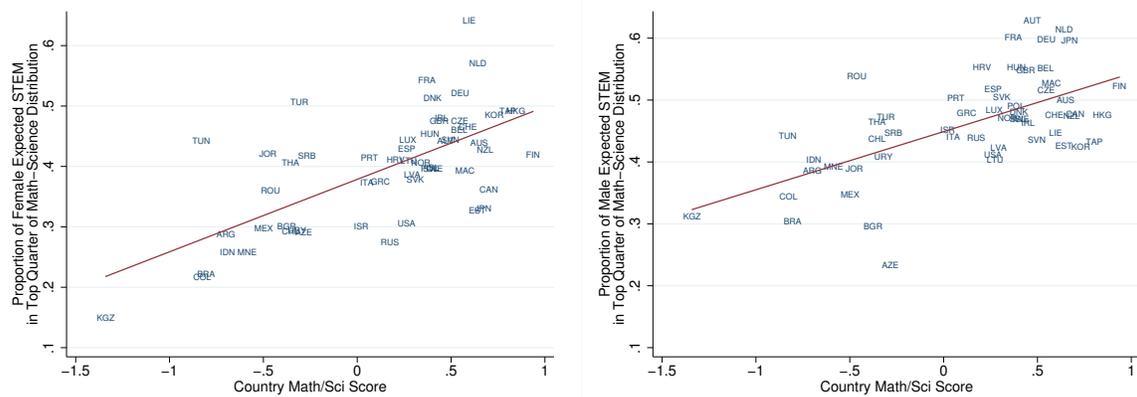
the GGI) into the models for the subset of countries for which we have scores to assess the possibility that ideals of self-expression explain the gender gaps in STEM aspirations. The pattern of coefficients suggests that the gender gap in STEM aspirations narrows when self-expression scores are higher, although the results are not statistically significant. The fact that the gap does not widen is consistent with the results we obtain using GGI. Finally, the coefficient on national tracking policy suggests that girls have relatively larger intercepts in countries that do not track their students before the age of 16 (net of other factors). This suggests that girls are more likely than boys to be influenced by the track or the signal that the track conveys in forming STEM aspirations.

We then performed three additional analyses using additional measures of STEM orientation and the national environment. The online supplement to this article contains the details of these results that we summarize here. First, we measured country differences in the national environments at other points than the mean of the distributions (cf. Penner 2008). The overall patterns are consistent with the predictions derived from expectation states theory: the main effects of country math-science performance and country self-assessment are smaller for girls than boys, but the returns for girls to their own math and science performance rise relative to the returns for boys in stronger math-science environments and in countries with higher math-science self-assessment. Second, although individual math-science self-assessment cannot be entered uncritically in a predictive model because it may be a consequence as much as a cause of STEM orientations, we elaborated our models to examine the potential consequence of taking individual self-assessment into account; the result was substantially similar to the models reported above. Third, we examined the consequences of restricting the dependent variable to the case of physical sciences only. The physical sciences model differs from the broader STEM model in various respects, but the pattern is entirely consistent with our theoretical expectations; specifically, the variation in the gender pattern in returns to own performance across national environments follows the same manner as when the broader STEM orientation outcome variable is used.

## Consequences for Gender Share and Gendered Ability Composition of Students with STEM Orientations

An indisputable consequence of the larger slopes on individual math-science performance in high-performing countries is that a higher share of STEM-oriented students will come from the top of the math-science distribution in those countries. Figure 7 shows the concentration effect for girls (in the first panel) and for boys (in the second panel).

To determine whether the concentrating effect of the performance context is stronger for girls than boys, we calculated the odds that a student would have a STEM orientation—first if the student had a math-science test score above the top quarter of the country math-science distribution and then if they had a test score below the top quarter—and then we calculated the ratio of the odds that such a student would be male relative to the odds that she would be female. The result

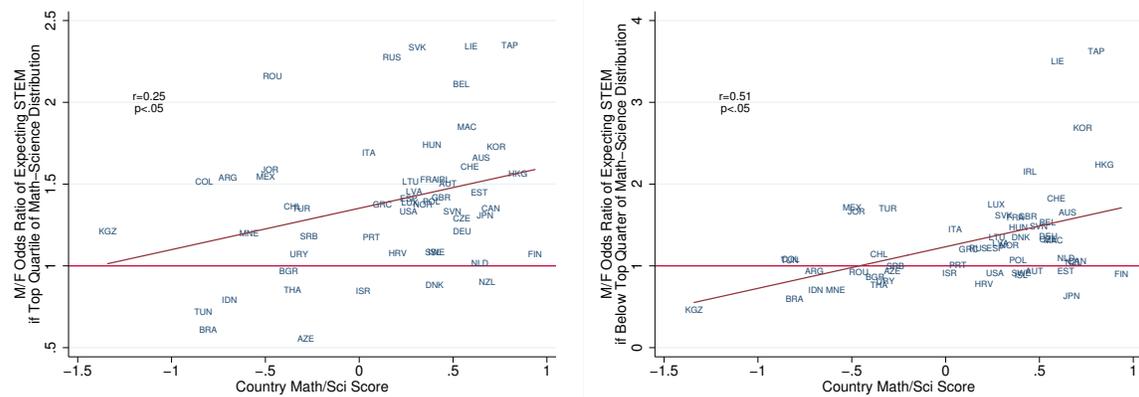


**Figure 7:** Proportion of male and female STEM-oriented students in the top quarter of the math-science distribution by national math-science performance level.

is presented in Figure 8. Figure 8 makes apparent that there is a greater gender difference in the concentrating effect of national performance environments for lower-performing students. The national performance environment is concentrating boys and girls with STEM orientations in the top of their math and science performance distributions to a lesser extent. Thus, stronger national performance environments increase the relative odds that boys in the bottom three quarters of the math-science distribution (compared to similar girls) will have STEM orientations. This macro-level difference between girls and boys is consistent with the predictions of expectation states theory.

We finally ask the question how the national performance environment relates to the relative share of students with STEM orientations who are female. The overall fraction of STEM-oriented students in a country depends on the intercept (the main effect of country, so to speak), the slope of math-science achievement, and the within-country distribution of performance among boys and girls. We have already seen that a stronger national performance environment increases the returns to math and science performance for girls even as it makes the intercept more negative. The combination of these effects leads to the relationship that is shown in Figure 9. We can see from this figure that the overall tendency is for countries with stronger math-science performance environments to have a disproportionately large number of STEM-oriented students who are boys.

To recapitulate, the national math-science performance environment generally drives down the overall fraction of STEM-oriented students who are female. Girls and boys respond differently to the national environment. Their overall rates of forming STEM orientations are depressed in stronger math-science performance environments, but the depressing effect is particularly strong for girls whose performance in math and science is relatively low. We see that the effects of expectation states theory manifest themselves in terms of cross-national differences in the pattern of forming self-assessments and in terms of the pattern of responding to own achievement.

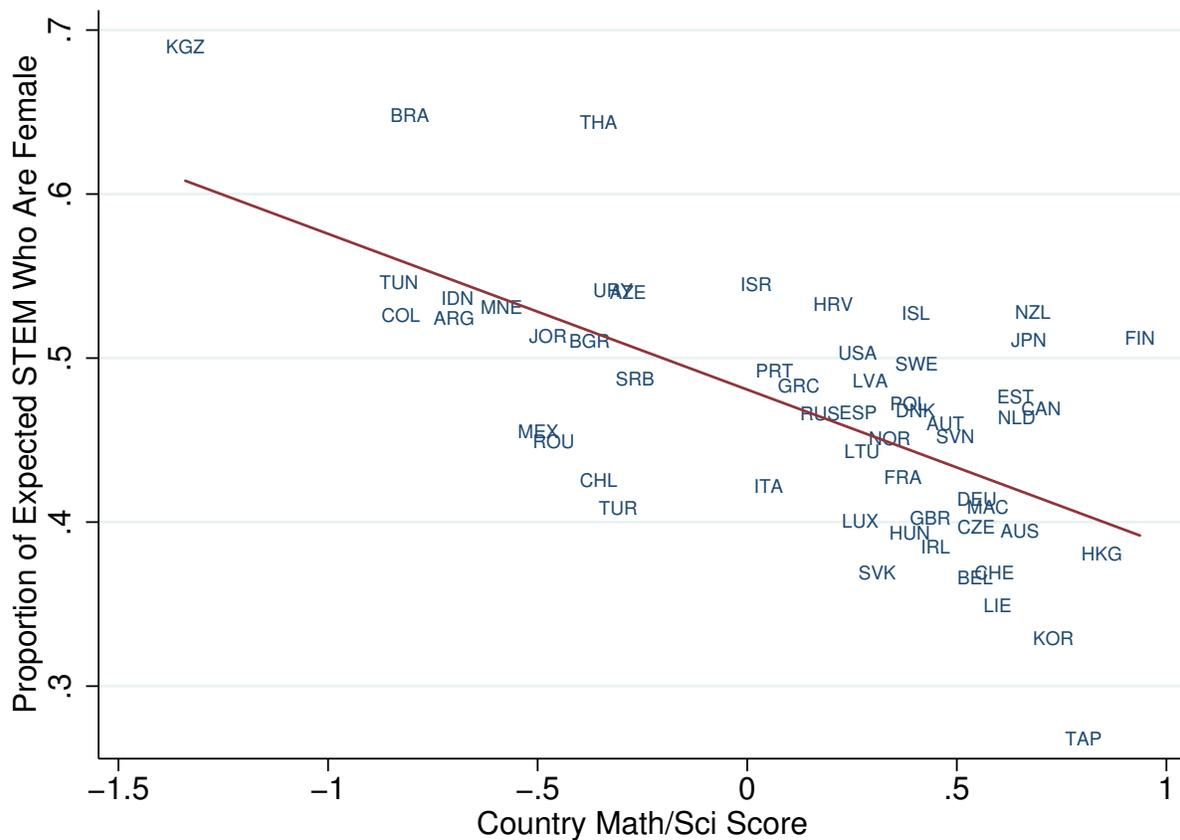


**Figure 8:** Odds ratio for male and female students that a student is STEM-oriented, given that the student is a top- or bottom-performing student in math and science, by the national math-science performance level (reference line represents even M–F odds).

## Discussion

This article makes three contributions. First, applying expectation states theory, we consider the impact that gender status roles have on gender differences in the formation of STEM orientations. We use a Bayesian decision-theoretic model to demonstrate how expectation states theory operates at the level of national environments. Status roles are represented by a country’s math and science performance level, the aggregate of its students’ performance-adjusted self-assessments in science, and the gender gap in the aggregated, performance-adjusted self-assessments in science. We treat nation-level variation in those dimensions as measures of the strength of the gender stereotyping regarding math and science in the culture. High performance in the environment implies more difficult curricula, higher performance standards, and greater competition, all of which have been shown to heighten gender-essentialist ideas about math and science. Higher country-level self-assessments are an aggregate of what Correll (2001) refers to as “biased” self-assessments; they suggest that students in a country require less evidence of their own abilities in order to form a STEM orientation. We demonstrate that higher-performance environments are characterized by lower country-level self-assessments and larger self-assessment gender gaps.

According to our theoretical model, girls and boys look to performance in math and science as a “signal” of suitability for STEM occupations. The stronger the math-science “signal” is, the higher the probability that he/she will adopt a STEM orientation. We support this hypothesis empirically by showing that around the world girls and boys are more likely to develop STEM orientations if they have stronger performance in math and science. In high-performance environments, we argue, the risk of failure increases for a given level of performance because the competition is strong. In such environments, boys and girls are likely to require stronger evidence that they are good in math and science before deciding to pursue a STEM orientation. We also support this empirically by showing that both boys and



**Figure 9:** Relationship between the proportion of girls among expected STEM students and the national math-science performance level.

girls have lower intercepts and stronger returns to math and science performance in stronger math and science environments.

The linchpin of the analysis, however, is our argument that prior beliefs about the competence of girls to do math and science both raise the cost of making the wrong choice if that choice is STEM and raise the standard that girls use to infer STEM suitability on the basis of academic performance. To that end, we show empirically that in countries with stronger math-science environments, the female intercept decreases and the female slope on own performance rises (relative to the male intercept and slope on performance). Similarly, in countries with stronger self-assessment effects, the male intercept increases and the male slope on own performance decreases (relative to the female intercept and slope on performance). Put differently, girls and boys respond differently to the environment, and their different responses are consistent with the predictions we derive from expectation states theory. We demonstrate that our findings are robust to several different specifications and sampling strategies.

Our second contribution relates to the literature that examines the persistent gender gap in the STEM work force. There are many factors that contribute to the lingering gap, but research in the United States suggests that a substantial contributing factor is the gap in STEM orientations in high school (Legewie and DiPrete 2014). In this article, we show that there are significant gender gaps in STEM occupational expectations around the world. We also show that the gender gaps in STEM expectations and attitudes towards math and science are largely unexplained by gender differences in math and science abilities. However, the consequence of that fact that girls hold themselves to a higher performance standard before forming STEM orientations in stronger math-science environments is that the overall gender gap in STEM aspirations increases. These effects are concentrated at the lower end of the performance distribution.

Our third contribution is to resolve the paradox in recent comparative literature on the gender gap in STEM that finds larger gender gaps in countries that score higher on indices of gender egalitarianism. We show that this paradoxical result comes from the fact that national indices of gender egalitarianism correlate positively with national performance in science and math, which produces a larger gender gap in STEM aspirations even as it raises the female slope on own performance relative to the male slope. Net of the powerful offsetting effects of the national performance environment, the gender gap in STEM orientations is generally smaller in countries with more gender-egalitarian cultures as measured by the gender gap index.

Importantly, the demonstration that, net of the performance environment, the STEM gender gaps shrink in countries with more gender-egalitarian cultures is robust to the precise theoretical interpretation that is given to the gender-specific effects of performance environments. In other words, the often-cited paradoxical apparently positive effect of gender egalitarianism on the gender STEM gap does not exist, and this is true regardless of whether the performance environment has its gender-specific impact by stimulating gender-essentialist ideas or whether it has this effect for some other reason. Previous explanations that seek to explain the paradox—through self-expression ideals or otherwise—are insufficient because they attempt to explain a paradox that does not exist, which calls to mind Merton's famous dictum about the importance of "establishing the phenomenon" in question as part of the process of "explain[ing] how they come to be" (Merton 1987:1,3). As we noted above, an extensive research literature supports the interpretation that high-performance environments stimulate gender-stereotyping of the performance task, which is why we favor this theoretical interpretation of our empirical results.

Our study has some limitations. Because it uses cross-sectional data, we can only examine associations between variables measured at one point in time. Not only does this preclude strong assertions about causality, but it also leaves open the question whether the national performance and self-assessment cultures on STEM orientation also influence STEM-related behaviors, such as course-taking and college majors in subsequent years. Nonetheless, our comparison of the relationship between self-assessment and STEM career plans in many countries provides a rich understanding of the cultural variation in the operation of gender stereotypes. A similar difficulty relates to the lack of definitive measures of country-level gender

status beliefs, as noted in the preceding paragraph. In the absence of a design that manipulates the extent of gender status beliefs while holding all else constant, we are unable to fully disentangle the impact of such beliefs from other national-level indicators that might correlate with gender status beliefs. A final difficulty is that we are unable to disentangle the separate contributions of math and science self-assessments on STEM career plans. We do show, however, that math and science achievements and math and science assessments are highly correlated, both at individual and country levels, as are the country-level gender gaps in those measures. As a result, it is unlikely that there are substantial differences in the contributions of math and science performance and self-assessments to the formation of STEM orientations.

## Notes

- 1 Lovaglia et al. (1998) differentiate self-assessments and expectations of competence that are not consciously held, but that still influence performance. In both cases, the emphasis is on the internalization of one's own status position, but status beliefs exert influences that extend beyond the effects on consciously held self-assessments.
- 2 Expectation states theory explains the emergence of relative performance expectations regarding group members that are formed when actors are oriented toward the accomplishment of a collective task or goal; status characteristics theory is an extension of expectation states theory that examines how cultural beliefs about status characteristics (including gender) organize the formation of expectation states (Correll and Ridgeway 2006). The theory has been expanded to recognize that status processes occur whenever situational demands pressure actors to evaluate their task performances relative to those of others (Correll 2001, 2004; Lovaglia et al. 1998). In an extension of expectation states theory, Foschi introduces "double standards" as a way in which differential performance expectations activate a different use of standards for attributing ability (Foschi 2000). Double standards theory argues that lower status members are evaluated by a stricter standards. This theory predicts that girls would demand a higher level of performance before forming STEM aspirations in cultures that have stronger gender stereotypes regarding math and science.
- 3 Most of the empirical evidence supporting expectation states theory comes from social psychological experiments (Berger et al. 1972; Correll and Ridgeway 2006). Although the experimental data provide important evidence regarding causality, they do not consider the impact of self-evaluations on the actual behaviors and aspirations of children embedded within schools or national environments.
- 4 The claims of competition playing a role in the decision dynamics need not imply that there is scarcity of opportunity for STEM positions, but rather that there is competition arising out of an enhanced performance standard. Countries with higher average achievement scores are likely to have more intensive mathematics curricula (Charles et al. 2014).
- 5 PISA 2012 excludes both the open-ended question that we use (see below) and the Likert statement "I would like to work in a career involving science." The only question on this topic asks a student to choose whether they plan to pursue a career "that involves a lot of mathematics" or a career "that involves a lot of science."
- 6 We use an open-ended question about occupational aspirations rather than a related question that uses a Likert-type scale to assess preferences for "careers involving science."

That question asks, “How much do you agree with the statements below? I would like to work in a career involving science,” and uses four agree–disagree response categories. The open-ended question is preferred because of the possibility of cultural bias in using the Likert item format for the dependent variable.

- 7 We constructed an alternate measure of the dependent variable that treated missing expected occupation data as non-STEM responses. Parallel analyses that use that measure confirm the patterns that we demonstrate below.
- 8 The findings are also robust to two other constraints—one that excludes the United States from the models altogether but retains reading test scores and one that excludes reading test scores from the models but retains the 2006 data for the United States.
- 9 Math and science test scores for individuals are highly correlated ( $r=0.91$ ,  $p<.05$ ). Math and science country effect sizes (Cohen’s  $d$ ), calculated as the difference in means in math for boys and girls and the difference in means in science for boys and girls, respectively, also are highly correlated ( $r=0.80$ ,  $p<.05$ ).
- 10 The PISA includes a socioeconomic measure called the index of Economic, Social and Cultural Status (ESCS), which is derived from the highest occupational status of parents, the highest educational level, and an index of household possessions. Immigrant status was derived from the question: “In what country were you and your parents born?” Either parent’s science-related career is derived from the questions “What is your mother’s main job?” and “What is your father’s main job?” These responses were coded using the International Standard Classification of Occupations, the same definition of STEM shown in the online supplement.
- 11 Information about the student’s study program was obtained through a student tracking form and the student questionnaire. All programs were classified using the international standard classification of education. We used the index for program orientation that was derived from the data on study programs.
- 12 The aggregate measures we create correlate strongly with similar measures in PISA 2003 and PISA 2012. To assess the strength of the relationships between math and science self-assessments, we reestimated the fixed-effects models using the 2003 and 2012 PISA data collections. In both data collections, the major content domain was mathematics. The gender gaps and country effects reported in this article correlate closely with effects computed using a similar specification with math self-concept as the dependent variable. The correlation coefficients are 0.67 for the gender gaps and 0.77 for the country effects (comparing the 2006 to the 2003 data) and 0.59 for the gender gaps and 0.75 for the country effects (comparing the 2006 to the 2012 data).
- 13 We also expect that math and science test scores will depend on the school and national context. However, it is not possible with these cross-sectional data to determine the extent to which context influences individual performance.
- 14 Setting all independent variables to zero corresponds at a substantive level to a 15-year-old, native-born student in the modal grade for the country, with average socioeconomic status, parents in non-STEM occupations, a placement in an academic track, and average values on all test-score measures.

## References

- Alon, Sigal, and Thomas A. DiPrete. 2014. Gender Differences in the Formation of Field of Study Choice Set. *Sociological Science* 2:50-81. <http://dx.doi.org/10.15195/v2.a5>

- Aronson, Joshua, Michael J. Lustina, Catherine Good, Kelli Keough, Claude M. Steele, and Baker, David P., and Deborah Perkins Jones. 1999. When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat. *Journal of Experimental Social Psychology* 35:29-46. <http://dx.doi.org/10.1006/jesp.1998.1371>
- Baker, David P. and Deborah Perkins Jones. 1993. Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance. *Sociology of Education* 66:91-103. <http://dx.doi.org/10.2307/2112795>
- Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. Status Characteristics and Social Interaction. *American Sociological Review* 37:241-255. <http://dx.doi.org/10.2307/2093465>
- Bönte, Werner. 2015. Gender Differences in Competitive Preferences: New Cross-Country Empirical Evidence. *Applied Economics Letters* 22:71-75. <http://dx.doi.org/10.1080/13504851.2014.927560>
- Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics* 129:1409-1447. <http://dx.doi.org/10.1093/qje/qju009>
- Cech, Erin A. 2013. The Self-Expressive Edge of Occupational Sex Segregation. *American Journal of Sociology* 119:747-789. <http://dx.doi.org/10.1086/673969>
- Charles, Maria. 2011. What Gender Is Science? *Contexts* 10:22-28. <http://dx.doi.org/10.1177/15366504211408795>
- Charles, Maria, and Karen Bradley. 2009. Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries. *American Journal of Sociology* 114:924-976. <http://dx.doi.org/10.1086/595942>
- Charles, Maria, Bridget Harr, Erin Cech, and Alexandra Hendley. 2014. Who Likes Math Where? Gender Differences in Eighth-Graders' Attitudes Around the World. *International Studies in Sociology of Education* 24:85-112. <http://dx.doi.org/10.1080/09620214.2014.895140>
- Cohen, Geoffrey L., and David K. Sherman. 2014. The Psychology of Change: Self-Affirmation and Social Psychological Intervention. *Annual Review of Psychology* 65:333-371. <http://dx.doi.org/10.1146/annurev-psych-010213-115137>
- Correll, S. J. 2001. Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology* 106:1691-1730. <http://dx.doi.org/10.1086/321299>
- Correll, Shelley J. 2004. Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review* 69:93-113. <http://dx.doi.org/10.1177/000312240406900106>
- Correll, Shelley J., and Cecilia L. Ridgeway. 2006. Expectation States Theory. In *Handbook of Social Psychology*, edited by John Delamater, *Handbooks of Sociology and Social Research* 29-51. Springer US. [http://dx.doi.org/10.1007/0-387-36921-x\\_2](http://dx.doi.org/10.1007/0-387-36921-x_2)
- Cotter, David, Joan M. Hermsen, and Reeve Vanneman. 2011. The End of the Gender Revolution? Gender Role Attitudes from 1977 to 2008. *American Journal of Sociology* 117:259-289. <http://dx.doi.org/10.1086/658853>
- Davis, James A. 1966. The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men. *American Journal of Sociology* 72:17-31. <http://dx.doi.org/10.1086/224257>
- Dey, Judy Goldberg, and Catherine Hill. 2007. *Behind the Pay Gap*. Washington, DC: AAUW Educational Foundation.

- DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools*. Russell Sage Foundation.
- Eccles, Jacquelynne S. 1994. Understanding Women's Educational and Occupational Choices. *Psychology of Women Quarterly* 18:585-609. <http://dx.doi.org/10.1111/j.1471-6402.1994.tb01049.x>
- Else-Quest, Nicole M., Janet Shibley Hyde, and Marcia C. Linn. 2010. Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin* 136:103-127. <http://dx.doi.org/10.1037/a0018053>
- Ferla, Johan, Martin Valcke, and Yonghong Cai. 2009. Academic Self-Efficacy and Academic Self-Concept: Reconsidering Structural Relationships. *Learning and Individual Differences* 19:499-505. <http://dx.doi.org/10.1016/j.lindif.2009.05.004>
- Foschi, Martha. 2000. Double Standards for Competence: Theory and Research. *Annual Review of Sociology* 26:21-42. <http://dx.doi.org/10.1146/annurev.soc.26.1.21>
- Foschi, Martha, and Ricardo Foschi. 1976. Evaluations and Expectations: A Bayesian Model. *Journal of Mathematical Sociology* 4:279-293. <http://dx.doi.org/10.1080/0022250X.1976.9989856>
- Fryer, Roland G., and Steven D. Levitt. 2010. An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics* 2:210-240. <http://dx.doi.org/10.1257/app.2.2.210>
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics* 118:1049-1074. <http://dx.doi.org/10.1162/00335530360698496>
- Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. Culture, Gender, and Math. *Science* 320:1164-1165. <http://dx.doi.org/10.1126/science.1154094>
- Hamamura, Takeshi. 2012. Power Distance Predicts Gender Differences in Math Performance Across Societies. *Social Psychological and Personality Science* 3:545-548. <http://dx.doi.org/10.1177/1948550611429191>
- Hansen, Stephen, Michael McMahon, and Carlos Velasco Rivera. 2014. Preferences or Private Assessments on a Monetary Policy Committee? *Journal of Monetary Economics* 67:16-32. <http://dx.doi.org/10.1016/j.jmoneco.2014.06.004>
- Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind*. McGraw-Hill, 3rd edition.
- Hyde, Janet S., and Janet E. Mertz. 2009. Gender, Culture, and Mathematics Performance. *Proceedings of the National Academy of Sciences* 106:8801-8807. <http://dx.doi.org/10.1073/pnas.0901265106>
- Iaryczower, Matias, and Matthew Shum. 2012. The Value of Information in the Court: Get It Right, Keep It Tight. *American Economic Review* 102:202-237. <http://dx.doi.org/10.1257/aer.102.1.202>
- Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge University Press. <http://dx.doi.org/10.1017/CB09780511790881>
- Kane, J. M., and J. E. Mertz. 2012. Debunking Myths about Gender and Mathematics Performance. *Notices of the AMS* 59:10-21. <http://dx.doi.org/10.1090/noti790>
- Legewie, Joscha, and Thomas DiPrete. 2014. Pathways to Science and Engineering Bachelor's for Men and Women. *Sociological Science* 1:41-48. <http://dx.doi.org/10.15195/v1.a4>

- Legewie, Joscha, and Thomas A. DiPrete. 2014. The High School Environment and the Gender Gap in Science and Engineering. *Sociology of Education* 87:259-280. <http://dx.doi.org/10.1177/0038040714547770>
- Lovaglia, Michael J., Jeffrey W. Lucas, Jeffrey A. Houser, R. Thye, Shane, and Barry Markovsky. 1998. Status Processes and Mental Ability Test Scores. *American Journal of Sociology* 104:195-228. <http://dx.doi.org/10.1086/210006>
- Mann, Allison, Joscha Legewie, and Thomas A. DiPrete. 2015. The Role of School Performance in Narrowing Gender Gaps in the Formation of STEM Aspirations: A Cross-National Study. *Frontiers in Psychology* 6:171. <http://dx.doi.org/10.3389/fpsyg.2015.00171>
- Marsh, Herbert W. 1987. The Big-Fish-Little-Pond Effect on Academic Self-Concept. *Journal of Educational Psychology* 79:280-295. <http://dx.doi.org/10.1037/0022-0663.79.3.280>
- Marsh, Herbert W., and Kit-Tai Hau. 2003. Big-Fish-Little-Pond Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools. *American Psychologist* 58:364-376. <http://dx.doi.org/10.1037/0003-066X.58.5.364>
- Marsh, Herbert W., and Alexander Seeshing Yeung. 1997. Coursework Selection: Relations to Academic Self-Concept and Achievement. *American Educational Research Journal* 34:691-720. <http://dx.doi.org/10.3102/00028312034004691>
- Merton, Robert K. 1987. Three Fragments from a Sociologist's Notebooks: Establishing the Phenomenon, Specialized Ignorance, and Strategic Research Materials. *Annual review of sociology* 13:1-29. <http://dx.doi.org/10.1146/annurev.so.13.080187.000245>
- Miller, David I., Alice H. Eagly, and Marcia C. Linn. 2014. Women's Representation in Science Predicts National Gender-Science Stereotypes: Evidence from 66 Nations. *Journal of Educational Psychology* 107:631-644. <http://dx.doi.org/10.1037/edu0000005>
- Nagengast, Benjamin, and Herbert W. Marsh. 2012. Big Fish in Little Ponds Aspire More: Mediation and Cross-Cultural Generalizability of School-Average Ability Effects on Self-Concept and Career Aspirations in Science. *Journal of Educational Psychology* 104:1033-1053. <http://dx.doi.org/10.1037/a0027697>
- Nagy, Gabriel, Jessica Garrett, Ulrich Trautwein, Kai S. Cortina, Jürgen Baumert, and Jacqueline S. Eccles. 2008. Gendered High School Course Selection as a Precursor of Gendered Careers: The Mediating Role of Self-Concept and Intrinsic Value. In *Gender and Occupational Outcomes: Longitudinal Assessments of Individual, Social, and Cultural Influences*, edited by H. M. G. Watt J. S. Eccles, pp. 115-143. Washington, DC, US: American Psychological Association. <http://dx.doi.org/10.1037/11706-004>
- Niederle, Muriel, and Lise Vesterlund. 2007. Do Women Shy Away from Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics* 122:1067-1101. <http://dx.doi.org/10.1162/qjec.122.3.1067>
- Niederle, Muriel, and Lise Vesterlund. 2010. Explaining the Gender Gap in Math Test Scores: The Role of Competition. *The Journal of Economic Perspectives* 24:129-144. <http://dx.doi.org/10.1257/jep.24.2.129>
- Nosek, Brian A, Frederick L Smyth, N. Sriram, Nicole M Lindner, Thierry Devos, Alfonso Ayala, Maliszewski, Félix Neto, Eero Olli, Jaihyun Park, Konrad Schnabel, Kimihiro Shiomura, Bogdan Tudor Tulbure, Reinout W Wiers, Mónika Somogyi, Nazar Akrami, Bo Ekehammar, Michelangelo Vianello, Mahzarin R Banaji, and Anthony G Greenwald. 2009. National Differences in Gender Science Stereotypes Predict National Sex Differences

- in Science and Math Achievement. *Proceedings of the National Academy of Sciences* 106:10593-10597. <http://dx.doi.org/10.1073/pnas.0809921106>
- Penner, Andrew M. 2008. Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors. *American Journal of Sociology*. 114:S138-S170. <http://dx.doi.org/10.1086/589252>
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. How Stereotypes Impair Women's Careers in Science. *Proceedings of the National Academy of Sciences* 11:4403-4408. <http://dx.doi.org/10.1073/pnas.1314788111>
- Ridgeway, Cecilia L. 2001. Gender, Status, and Leadership. *Journal of Social Issues* 57:637-655. <http://dx.doi.org/10.1111/0022-4537.00233>
- Ridgeway, Cecilia L. 2011. *Framed By Gender: How Gender Inequality Persists In the Modern World*. Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199755776.001.0001>
- Ridgeway, Cecilia L., and Shelley J. Correll. 2004. Unpacking the Gender System A Theoretical Perspective on Gender Beliefs and Social Relations. *Gender & Society* 18:510-531. <http://dx.doi.org/10.1177/0891243204265269>
- Riegle-Crumb, Catherine, and Melissa Humphries. 2012. Exploring Bias in Math Teachers' Perceptions of Students' Ability by Gender and Race/Ethnicity. *Gender & Society* 26:290-322. <http://dx.doi.org/10.1177/0891243211434614>
- Sikora, Joanna, and Artur Pokropek. 2012. Gender Segregation of Adolescent Science Career Plans in 50 Countries. *Science Education* 96:234-264. <http://dx.doi.org/10.1002/sce.20479>
- Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology* 35:4-28. <http://dx.doi.org/10.1006/jesp.1998.1373>
- Steele, Claude M. 1997. A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist* 52:613-629. <http://dx.doi.org/10.1037/0003-066X.52.6.613>
- Voyer, Daniel, and Susan D. Voyer. 2014. Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin* 140:1174-1204. <http://dx.doi.org/10.1037/a0036620>

**Acknowledgements:** This project was supported by Award Number R01EB010584 from the National Institute of Biomedical Imaging and Bioengineering. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Biomedical Imaging and Bioengineering or the National Institutes of Health. We acknowledge helpful comments by Claudia Buchmann and Joscha Legewie.

**Allison Mann:** Sociology Department, Columbia University.  
E-mail: alm2174@columbia.edu.

**Thomas A. DiPrete:** Sociology Department, Columbia University.  
E-mail: tad61@columbia.edu.