

This is a pre-print of an article published in [insert journal title]. The final authenticated version is available online at: <https://doi.org/10.1007/s10648-019-09493-1>

THE INTERSECTION OF GENDER, SOCIAL CLASS, AND CULTURAL CONTEXT: A META-ANALYSIS

Philip D. Parker^{1*}, Brooke Van Zanden¹, Herbert W. Marsh¹, Katherine Owen², Jasper J. Duineveld¹ & Michael Noetel¹.

¹Australian Catholic University, Institute for Positive Psychology and Education

² University of Sydney, SPRINTER Research Group and Prevention Research Collaboration

*Corresponding author.

Professor Philip Parker

Deputy Director and Program Lead, Institute for Positive Psychology and Education

Australian Catholic University

T: [+61 422 890 578](tel:+61422890578)

E: philip.parker@acu.edu.au

W: www.acu.edu.au

ORCID: [0000-0002-4604-8566](https://orcid.org/0000-0002-4604-8566)

Acknowledgements: This study was funded by the Australian Research Council (DP130102713).

Data and supplementary material available at <https://osf.io/mcavt/>

Abstract

Expectancy Value Theory is often evoked by educational psychologists to explain gender differences in Science, Technology, Engineering, and Mathematics (STEM) variables. Yet gender does not operate in isolation. Nor are gender effects likely to be context free. In the current meta-analysis, we explore gender differences in STEM related expectancy for success, and the task values of intrinsic, utility, attainment, and cost. We find gender differences were generally small in size. Invoking the concept of intersectionality, we find that gender differences are moderated, primarily, by socio-economic status, ethnic diversity and national gender equality. Higher socio-economic status and more gender equality appear to be related to greater gender differences.

Keywords: Gender Differences, Expectancy Value Theory, STEM, Intersectionality

The Intersection of Gender, Social Class, and Cultural Context: A Meta-Analysis of Self-Concept and Task Value

The number of students enrolling in senior high school science and mathematics courses is on the decline, and this decrease appears most prominent among women (Kennedy, Lyons, & Quinn, 2014; Mack & Wilson, 2015; Office of the Chief Scientist, 2014). Consequently, there has been a strong push from governments of post-industrial nations to encourage the retention of women in science education and careers (e.g., National Innovation and Science Agenda, 2017; National Science and Technology Council, 2013). Eccles and colleagues have focused on the role of expectancy value theory (EVT) as a means of explaining gender differences in STEM enrolment (see Eccles, 1994; Eccles, 2005; Eccles & Wigfield, 2002). EVT proposes that the central determining factors of whether a person will attempt a given action is whether they believe success is likely and they value the anticipated outcomes of that action. In our context, EVT proposes that a girl will enroll in a STEM course if she thinks that she can succeed (expectancy) and she sees success in that field as valuable (value). Eccles and colleagues' highlight the role of social and cultural context in determining expectancy and value. Eccles and colleagues (e.g., Eccles & Hoffman, 1984; Eccles & Jacobs, 1986; Eccles, Jacobs, & Harold, 1990) argue that gender differences in educational and career pathways are primarily the result of gender socialization experiences that affects young people's self-beliefs and values. This different choice architecture leads to the different career choice patterns of boys and girls. According to Eccles (1994), girls disengage from math and science, not because they lack ability or aptitude, but because of low expectancies for success and low task value in STEM subjects that are in large part shaped by social and cultural milieu.

Our focus in this meta-analysis is on gender differences in expectancy and task value variables for science, math, and technology—juxtaposed against verbal academic domains. Expectancy is defined as an individual's belief about their ability to perform a given task or set of tasks (Eccles & Wigfield, 2002). Theoretically expectancy is different from self-concept in Eccles EVT theory. Empirically, however, researchers have determined that the two variables are so closely related that they are nearly synonymous (Eccles & Wigfield, 2002). As such almost all research in this area operationalizes expectancy as self-concept (e.g., Guo et al., 2015a, 2015b; Nagengast et al., 2011; Trautwein et al., 2012). Expectancies are about whether a person believes they can perform a given task. Task-value is about whether that individual ascribes any personal importance to that task (Eccles & Wigfield, 2002).

Eccles EVT model argues there are four components of task value: attainment value (the importance ascribed to doing well in a task), intrinsic value (enjoyment that an individual gets out of doing a task), utility value (the instrumental value that a task have for future plans), and cost (the opportunity cost that a task imposes). These constructs are important because they are thought to underlie both educational and occupational choice. Thus, they are critical in understanding gender differences in STEM course enrolment—in both high school and college—and STEM occupational attainment (Lauermann, Tsai, & Eccles, 2017). Distinguishing different components of value is important because they capture both extrinsic and intrinsic reasons that explain why people make the achievement related choices that they do (Wigfield & Eccles, 2000). The more capable a person feels, the greater value—particularly intrinsic value—they assign to a particular task, and the lower the costs associated with engaging in that task, the more likely a person is to choose to engage in it. Thus, EVT variables are not only powerful explanatory variables, they also represent clear intervention targets that can be used to address gender differences in STEM choices.

As we outline below, we apply an intersectional approach to reviewing the literature on gender differences in expectancy and value constructs. Foreshadowing our discussion of this topic an intersectional approach is not merely interested in determining the relative size of gender differences in these variables, but in exploring the heterogeneity in these effect sizes and what social categories and other factors may explain such heterogeneity. Thus, consistent with the quantitative intersectional perspective of Else-Quest and Hyde (2016b) in our meta-analysis we are primarily interested in determining whether gender differences in expectancy and value are moderated by participants ethnicity, age, and social class. In this paper, we also test the relationship between gender differences in STEM variables and country level gender inequality. Labelled the gender stratification hypothesis, Else-Quest, Hyde, and Linn (2010) hypothesized that gender differences would be smaller in countries where opportunity structures were better for girls. Recently, the gender stratification hypothesis was tested in Australian age cohorts born from 1981 to 1993—under the assumption that gender inequality has lessened somewhat over time—and found partially supported (Parker, Van Zanden, & Parker, 2018). As such we also include publication era as a moderator.

EVT and Context

Eccles and Wigfield (2002, pg. 128) state that it is “difficult if not impossible to understand students’ motivation without understanding the contexts they are experiencing”. Despite the wealth of research on gender differences and similarities in educational and occupational outcomes, there has been comparatively less research that has investigated the specific contexts in which gender differences in educational outcomes may vary. Indeed, this research gap is reflective of a wider problem within psychology, whereby psychology’s focus on the individual in isolation to their contexts means that the impact of societal and structural inequality on marginalized individuals is often ignored (Fox, Prilleltensky, & Austin, 2009). If we fail to consider how social, cultural, and political contexts of individuals feed into thoughts, appraisals, wellbeing, and values, then we are neglecting a potentially powerful and important influence on people’s inner lives and experiences. A lack of within-study contextual moderators is often a barrier to understanding contextual influences. This is a problem that may be overcome by exploring between study variability in results that differ in the context in which the research was undertaken. Thus, the current meta-analysis aims to help rectify this situation by examining the available literature on gender differences in EVT constructs, determine what contextual factors moderate these differences, and identify where research evidence is most sparse.

The Intersection of Gender and Socio-cultural Contexts

Psychological research on gender has also faced difficulties in incorporating context. Criticism has been levelled at psychology over the lack of investigation into how gender intersects with other social and cultural categories (Else-Quest & Hyde, 2016a; Hyde, 2007). Eagly, Eaton, Rose, Riger, and McHugh’s (2012) review on feminism and psychology found that only a small minority of studies on the psychology of women and gender have attended to the heterogeneity or diversity within gender by including analyses of gender across social class, sexual orientation and ethnicity. Eagly and colleagues noted that the intersection of gender with social class and sexual orientation was particularly under-researched, and thus flagged intersectionality as an area needing more research. Similarly, Hyde (2007) emphasized that much of what we currently know about gender and psychology is actually the experience of gender for American middle-class white college students (Hyde, 2007; see also Henrich, Heine, & Norenzayan, 2010). Consequently, little psychological research has investigated how gender effects in EVT are moderated by ethnicity, class, and nationality. Hyde (2012; 2013) flags this as an area in which new research is

crucial, investigating not only gender, but also the *intersection* of gender with social categories like class and ethnicity.

To explore this intersection, we need data from many social categories to see if the relationship between gender and EVT variables changes in those categories. But we also need to direct attention to heterogeneity in effect sizes more generally as quantitative analysis may struggle to adequately capture the myriad of intersections that influence women's and men's self-beliefs and values. Attention to heterogeneity and where possible identifying critical intersections will provide evidence that may help better target interventions, determine how research resources could be better deployed, and potentially identify mechanisms that can help understand how gender differences in EVT variables emerge.

Are Intersectional Research Questions Compatible with EVT and Quantitative Research?

Intersectional feminism has its roots in the Black feminism of Hull, Scott, & Smith (1982)—in the book *All the women are White, all the men are Black, but some of us are brave*—but was formalized by Crenshaw (1989) and Collins (1990). Intersectional feminism emerged in part from critical theory that challenged positivist assumptions about science and knowledge (see Bowleg, 2008 for a critique of quantitative methods in relation to intersectionality). Indeed, some suggest that intersectionality may lead to poor research if not framed within a critical theory ontology and epistemology (Hancock, 2016, p.13). Our research largely relies on quantitative methods to explore questions about gender and educational attainment in the context of EVT. If intersectionality must rely on a critical theory orientation such research would seem difficult. But this need not be the case.

Else-Quest and Hyde (2016a) state that intersectionality can use a critical empirical, social constructivist, or standpoint epistemology. Martinez, Martin, and Marlow (2014) have advanced a critical realist approach in contrast to critical theory perspectives. Likewise, Muntaner and Augustinavicius (2019) have recently provided an account of intersectionality from the perspective of Scientific Realism. This ontological and epistemological freedom is partly due to the myriad of ways in which intersectionality has been used. Martinez et al. (2014, p. 449) states that intersectionality “has been considered a theory, a paradigm, a framework, a method, a perspective, or a lens”. Else-Quest and Hyde (2016a) state that intersectionality is any research where: a) there is a recognition that everybody is defined by multiple interconnected and enmeshed social categories; b) that power and inequality are embedded in social categories and the relationships between them; and c) that these categories are properties of multiple levels of person and context and that these social categories are thus fluid and dynamic. Following these criteria, we take a

critical realist approach (see Collier, 1994 for a review). A critical realist approach to intersectionality considers categories like gender to be an abstraction that, although not capturing the reality of a particular person, nevertheless represents a structural position with political, social, economic, and cultural contexts (see Gunnarsson, 2011). In such an approach, categories like gender are useful objects of analysis particularly when considering the intersection of gender with other social categories, as well as with temporal, cultural, political, and social contexts, that may moderate the universal abstraction of gender (Martinez et al., 2014). We argue our position frames both abstracted social categories and heterogeneity within these categories as valuable foci of research.

Consistent with Else-Quest and Hyde's (2016b) we argue that quantitative methods can provide insight into both gender as a universal category and provide insight into heterogeneity within it. Crucially, Else-Quest and Hyde (2016b) highlight meta-analyses of gender differences that incorporate moderators as one quantitative method that can be used from an intersectional perspective. Meta-analysis is useful because populations from which data are drawn are typically more diverse in meta-analysis than would be the case in a single study. Thus, meta-analysis can pay particular attention to both universal gender effects and heterogeneity in these effects. Further, moderators can be easily incorporated to seek to explain this heterogeneity in terms of other social categories or other contextual variables.

While meta-analysis is an extremely useful tool for intersectional research, our critical realist approach demands that we scrutinize the methods limitations. Crucially, researchers who use meta-analysis must rely upon the definitions and thus the operationalizations of core constructs of others research. For us, some constructs we explore may be defined narrowly or via limited information. For example, we explore gender differences between boys and girls only. This is because so little quantitative research in educational psychology provides information beyond these binaries. This is a concern as, for example, estimates suggest the intersex individuals could make up a not insignificant proportion of the population (Parker, Larkin, & Cockburn, 2017). Further, research uses a wide variety of ways of defining socioeconomic status or social class, and indeed does not distinguish between the two. As we have outlined in our previous research (Author et al., 2017) social class and socioeconomic status are highly related but different concepts as is evident by their prediction of different outcomes. Yet here we must rely on researchers' assessments of social class, generally expressed in summative statements such as "most of the sample came from a working-class background". Mostly the researcher's definition of social class is based on parental occupational prestige, earnings capacity, and, typical in US research, qualification for free or reduced-price lunches. Such

definitions are generally crude. Class definitions in this area largely ignore forms of capital like social and cultural. And descriptions of the study sample generally conflate social-class and socio-economic status. Such concerns also relate to the way in which researchers describe and define ethnicity. Such limitations are important, but the wealth of information that comes from meta-analyses and the ability to focus so strongly on heterogeneity in gender differences and what moderators may explain it mean that our endeavors are valuable. Further, given we treat social categories as abstractions that indicate structural positions mean that using the typical thin definitions of categories presented in the literature is not only justifiable but informative about how those categories are perceived by the research community.

A Narrative Review of the Literature

The intersection of gender, ethnicity and cultural difference has sparked the interest of a number of quantitative researchers in recent years. This is not surprising given the strong traditions of intersectional research and feminists who have examined relations between sexism and racism. However, there has been less research on how gender intersects with other important social categories (Eagly et al., 2012). For instance, how does social class affect the experience of being a woman? Or, how does the experience of gender compare across different geographies (e.g., Is the effect of gender on EVT-related variables different for women in high density urban cities, versus women in rural or remote areas)? We review the existing literature here.

Social class, gender and math outcomes. There have been only a handful of studies that have included analyses of the interaction between gender and social class with relation to math achievement, and almost none in relation to EVT-related attitudes. Studies have shown that the small gender achievement gap in math becomes larger for American students from higher SES backgrounds, and that this effect occurred from elementary to high school (Lubienski, Crane, and Robinson, 2011; McGraw et al., 2006). Interestingly, McGraw et al. (2006) found that this pattern only extended to white rich students, with no evidence of the same trend amongst students from other ethnicities. Similarly, Fryer and Levitt (2010) found a gender difference in declining math performance during adolescence was largest for young girls from the highest quintile of social class, girls with highly educated mothers, and girls who attended private schools.

Why is the gender gap in math performance larger for richer people? One reason could be that children from wealthier, resource-rich families are provided with more gender socialization opportunities. Indeed, there is evidence to suggest that children from higher socio-economic statuses are exposed to more gender specific parenting

patterns and extra-curricular activities outside of school, compared to their peers from lower SES families (Lareau, 2003; Lubienski, Robinson, Crane, & Ganley, 2013). Research from Sáinz and López-Sáez (2010) on computer attitudes and behavior showed preliminary evidence for this mechanism using self-report data from Spanish adolescents. Sáinz and López-Sáez found that the magnitude of the gender gap grew larger in higher SES adolescents, however, this interaction effect did not replicate for the affective measure of computer attitudes used in the study.

Experimental research using vignette that varied child gender and social class show teachers are influenced in their assessments by these social categories (Auwarter & Aruguete, 2008a). Auwarter and Aruguete's study featured a student who was struggling at school and failing math despite having an IQ that was not below average. Results showed that teachers were more likely to judge the personal characteristics (e.g., competence in math) of low SES girls favorably compared to their high SES counterparts, who received the harshest evaluations of their personal characteristics out of any demographic group. In contrast, low SES boys received more negative ratings in comparison to high SES boys—who received the most favorable ratings of any demographic group. Furthermore, high SES girls were the least likely demographic to be referred to academic support or assistance (e.g., math tutoring).

Locale, gender and math outcomes. Quantitative research on how gender and locale might interact to influence expectancy and value is rare. In a study of gender differences in cognitive abilities amongst Peruvian children, Stevenson, Chen, and Booth (1990) found that for math achievement, there was a trend of an increasing gender gap in urban areas compared to rural areas. For example, young unschooled or first grade girls from the urban area of Lima experienced greater disadvantage relative to unschooled or first grade boys from Lima ($d = 0.26$). In contrast, gender effects for girls and boys from more rural areas of Peru were ($d = 0.05$ and 0.09) amongst non-schooled and first grade children. This trend extended to first, second and third grade older children to some degree ($d = 0.29$ in urban area, compared to ($d = 0.15$ and -0.07 in rural areas). Sáinz and López-Sáez (2010) showed similar results in the area of computer studies, finding that the gender gap in the behavioral dimension of computing attitudes (i.e., time spent using computers) was larger for urban adolescents, and smaller amongst rural adolescents.

Ethnicity. Out of the thousands of studies on attitudes towards science and math, only a small number have focused on the intersection of gender and ethnicity, and most research has failed to include ethnically diverse

samples in research on the STEM gender gap. Catsambis (1994; 1995) was one of the first researchers to explore this research question with regards to math and science attitudes. Catsambis found that there was some variability in gender differences across Black, Latino/Latina and White students. White students had the largest gender gaps for science attitudes, and Latino/Latina students exhibited the largest gender gaps for math attitudes.

Other researchers have explored gender and ethnicity with regards to technology in addition to math. Zarrett, Malanchuk, Davis-Kean, and Eccles (2006) found gender differences favoring boys in math and technology attitudes that existed regardless of whether students were African American or White. More specifically, gender differences in computer self-concept for programming and developing software, were moderate for White students and small for African American students. Gender differences were almost non-existent for White students in computer self-concept in word processing and accounting, and very small for African American students (favoring girls). Data about intentions to enroll in computer science majors show a similar pattern, whereby the gender gap for white students is (9:1), compared to (4.5:1) for other Latino, African American and Asian American students (NSF, Division of Science Resource Statistics, 2012). These findings suggest that gender gaps are largest between White boys and girls.

A similar pattern can be observed for math and science achievement, where the gender gap is greatest amongst White students (Catsambis 1994; 1995, Coley, 2001; Else-Quest et al., 2013; McGraw, Lubienski, & Strutchens, 2006). Thus, there seems to be some evidence of an ethnicity by gender interaction amongst American samples, whereby gender gaps are the largest and most consistent for White and Latino/Latina students, and smaller for Asian American and African American students.

However, a recent study of adolescents by Else-Quest et al. (2013) has demonstrated that while there were some slight variations in the size of gender differences between Latino/Latina, Asian American, and Caucasian students for self-beliefs and values in science and math, these differences tend not to be statistically significant. While these studies have provided valuable preliminary insights into the exploration of the intersection of gender and ethnicity for STEM attitudes, there has been little attempt to explore the ethnicity and gender interaction in settings outside the US.

National-Level Indicators of Gender Equality and Relations to EVT variables. Social and structural inequality is not merely a political issue but is critical to individual functioning and wellbeing (Glick & Fiske, 2001; Jenkins, 2000; Lykes, 2000; Prilleltensky, 2008; Zurbriggen & Capdevila, 2010). One way to investigate the

relationship between macro structures of power and inequality is to assess the relationships between individual level thoughts, feelings, and behaviors, and national level indicators of social equality. In this vein, the gender stratification hypothesis argues that gender differences in STEM achievement variables are due to differences in the political, economic, or cultural context that structure individual's choice architecture (Else-Quest et al., 2010). Empirical research in this area is growing (e.g., Baker & Jones, 1993; Else-Quest, Hyde, & Linn, 2010; Fryer & Levitt, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Guo et al., 2019; Hyde & Mertz, 2009; Kane & Mertz, 2012; Penner, 2008). Yet the theory has found some support (e.g., Else-Quest et al., 2010), no support (e.g., Ireson, 2017), or paradoxically suggest that gender differences increase with greater gender equality (e.g., Stoet & Geary, 2018). This erratic literature may be due to the different ways in which global gender equality is defined; as either composite (relating to the general treatment of women) or domain specific (relating to equality in indices directly related to STEM).

Else-Quest and colleagues (2010) analyzed the data from PISA and TIMSS with a series of global and domain specific measures. They found that when equality was measured with domain specific measures of gender equality, gender differences in EVT variables declined—thus supporting the gender stratification hypothesis. The authors also claim that composite measures of gender equality produced few clear results. Given the growing interest in the relationship between national indicators of gender equality and gender differences we explore both the composite and domain specific measures of gender equality used in Else-Quest et al. (2010) in order to explore macro-context as a potential moderator of gender differences in EVT variables.

Current Research

The central aim of this study is to address the dearth of research exploring gender differences in relation to social class and cultural contexts and to do so a meta-analysis using between study variation to explore such moderators. Hyde (2013) notes that there are few meta-analyses on gender that have tested gender differences across other social categories such as ethnicity and class. A meta-analysis, given that it can draw upon data from multiple populations, can also consider more macro-contextual variables that may structure the position of women differently. Meta-analytic approaches to intersectionality provides an important opportunity both to estimate average estimates of gender differences in EVT variables and to explore heterogeneity and what factors may help explain it.

This is of course dependent on researchers adequately describing their sample. Hyde (2013) states that this has not traditionally been standard practice in the field. Thus, not only do we consider between study variance in samples, we also consider the degree to which studies fail to sufficiently describe their sample to allow for such research.

Aims, Research Questions, and Hypotheses

Aim 1. We aimed to describe the state of the literature on gender differences in EVT research.

Research Question 1.1: What proportion of studies explain their sample in sufficient detail to extract the potentially important moderators – socio-economic status and ethnic diversity – of the gender difference in EVT?

Research Question 1.2: What are the current gaps (e.g., understudied constructs or academic domains) in the literature on gender differences across expectancy for success and task value variables?

Aim 2. We aimed to determine the average size of gender differences in EVT constructs.

Hypothesis 2.1: Gender differences will follow a gender stereotypical pattern, whereby math and physical science constructs will exhibit the largest gender differences favoring boys, while verbal domains and biological science constructs will exhibit the largest gender differences favoring girls.

Hypothesis 2.2: Although gender stereotypical patterns will emerge, the pooled effect size for each academic domain will be within the small-moderate range.

Aim 3. We aimed to determine the degree of heterogeneity in effect sizes for gender differences in EVT constructs and to see if we could explain this heterogeneity through within and between study moderators.

Research Question 3.1: How much heterogeneity (study-to-study variation) exists across effect sizes?

Research Question 3.2: What moderators help explain this heterogeneity (e.g., percentage of sample from an ethnic minority and social class of sample)?

Research Question 3.3: What moderators exist at the country level?

Hypothesis 3.1: Consistent with several strands of evidence reviewed above, gender differences in STEM variables will be larger in countries with higher level of gender equality as measured by composite measures but gender differences will be smaller in countries with greater gender equality when domain specific measures are used.

Method

Eligibility criteria

To be included in this review, studies were required to report a quantitative relationship between gender and a measure of domain specific (e.g., math, science) expectancy for success and domain specific task value. All studies were required to have full-text English results to meet eligibility criteria to ensure that the data extracted were accurate and representative of the study in question. Effect sizes from meta-analyses and other review articles were excluded. There were no restrictions on publication date or type, participant age or any other demographic factors (see supplementary material for a full list of the eligibility criteria used in this study).

Information sources

Searches were conducted within Psychinfo, Education Resources Information Center (ERIC), and Web of Science. Combinations of key words relating EVT terms were used to identify eligible studies in April 2016. The search strategy is available in the supplementary material. Both peer reviewed and grey literature were extracted

Search

The search strategy aimed to encompass a variety of terms and concepts that tapped into the constructs of expectancy for success and value. Database specific search strategies are available in the supplementary material. However, the general search string was:

(math* OR "verbal ability" OR English OR Science OR STEM) **AND** (gender OR sex) **AND** ("self-concept" OR expectancy OR "expectancy for success" OR "math self-concept" OR "verbal self-concept" OR "English self-concept" OR "science self-concept" OR "self-efficacy" OR competenc* OR attitude OR "expectancy value theory" OR "EVT") **AND** ("task value" OR value OR interest OR "intrinsic motivation" OR "intrinsic value" OR enjoyment OR importance OR "attainment value" OR usefulness OR "incentive value" OR "utility value" OR "extrinsic motivation")

Readers will notice that this strategy limits our search to papers that include at least on self-concept and at least one task value construct. In the vast majority of studies there is little or no assurance that the samples are representative of any population. However, at least for studies that include both an expectancy and a value there is an inherent control for the sample that facilitates comparison of these major dimensions. Hence, this strategy facilitates the comparison of these two major constructs.

Study Selection

All potentially eligible studies were exported into a single Endnote library where duplicate studies were removed. Next, each record was independently screened by at least two authors (BV, MN, KO) Records were excluded when both reviewers determined that the record did not meet eligibility criteria. Finally, full-text versions of the remaining articles were obtained and independently screened for eligibility. Discrepancies regarding inclusion were resolved by discussion between the researchers.

Data Collection Process

Three researchers extracted the data from eligible studies. Extracted data included the year of publication, gender split of participants, the domains in which expectancy/value was measured, country in which the study was conducted, socio-economic status of the sample, ethnic minority percentage of the sample, expectancy and value measures used, mean age, and the statistical result that examined the effect of gender on expectancy and value.

Summary Measures

We used standardized mean differences, correlation coefficients, t values, and f values to calculate the Cohen's d effect sizes for each study. All summary measures were converted to Cohen's d using Rosenthal's (1991) and (1994) conversion formulas. Effect sizes (d) were reported in keeping with Cohen's (1988) general guidelines for interpreting effect sizes; .2 (small), .5 (medium), and .8 (large). All effect sizes were calculated such that girls' scores on the EVT variables were subtracted from boys' scores such that a positive number indicates an advantage for boys.

Analysis

Traditional meta-analyses have used fixed and random effects models to analyze data. However, these approaches are limited in that they assume independence (Field, 2003; Author, 2009), meaning that only one effect size per study can be included in the meta-analysis. Traditional methods of dealing with this (e.g., average effect sizes, or reporting only one effect from a study) are problematic in that they can lose vital information, and limit the testing of moderators (Cheung, 2014).

One way to overcome these challenges is the utilization of structural equation modelling and multilevel modelling approaches to meta-analysis (Goldstein, 1995; Author, 2009; Raudenbush & Bryk, 1985, Van Den Noortgate & Onghena, 2003). Thus, in this meta-analysis we took a structural equation approach to multi-level meta-analysis. We conducted all analyses in R using the package metaSEM (Cheung, 2011), using unconditional mixed-

effects models to calculate overall pooled effect sizes (pooled d) and their respective 95% confidence intervals (CIs). Significant effects were reported when the 95% CIs did not cross zero.

To test heterogeneity in pooled effect sizes, we used the I^2 statistic (Higgins, Thompson, Deeks, & Altman, 2003). When effect sizes were heterogeneous (for example an I^2 that was above 25%), moderator analyses were conducted to explore the degree to which study and sample characteristics could explain heterogeneity in the pooled effect size. For each moderation analysis, we reported the proportion of explained variance of heterogeneity that can be accounted for by the inclusion of a moderator variable (R^2), and the heterogeneity between effect sizes in each category (I^2). However, one complication of this was that typically at least 4 effect sizes are required in each moderator sub-category in order to calculate accurate results (Fu et al., 2011). Thus, we only included moderation analyses on variables that had enough data to reach reliable conclusions.

Moderators

Moderators included social class of samples (e.g., working class, middle-class, and upper class majority sample in each study), percentage of participants belonging to an ethnic minority within a study¹, average age of study participants (elementary school, middle school, high school, young adult, and adult), publication date, population type (e.g., advanced or elective student populations), publication type (e.g., peer-reviewed versus theses) and the internal consistency of the scale used to measure a given EVT variable (>.70 versus <.70). We attempted to collect data at the within study level where possible (e.g., separate effect sizes for gender for a different age group within a study), however, this was dependent on whether studies reported individual effect sizes for subgroups within a study.

For country level indicators, we used both composite measures (i.e., measures that aimed to provide an overall view of gender equality in a given context) and domain specific measures (measures of gender equality that reflect relative standing in education and occupation) as a means of replicating Else-Quest et al. (2010). In line with recent advances in thinking about these measures (see Else-Quest & Hamilton, 2018), we focused our attention on the composite measures of the Global Gap Index (GGI; taken from The World Bank www.tcddata360.worldbank.org) and the Gender Inequality Index (GII; taken from the Human Development Reports <http://hdr.undp.org>). Domain specific measures were extracted from world bank data or from the human development reports including the primary enrolment ratio of girls to boys (Prim.), secondary enrolment gender ratio (Sec.), tertiary enrolment gender ratio (Tert.), women to men labor force participation ratio (LP), women's share of

research positions (WR), and women's share of parliamentary seats (Parl.). We used the most recent year of all indexes (2017 for GII and 2018 for all others). Not all indexes were available in all countries for the most recent year. When a 2017/2018 score was not available for a given country, we used the most recent year where data was available. We also recoded the indexes such that high scores always equaled greater gender equality (e.g., the GGI was reverse coded).

Publication Bias

Funnel plots were examined to assess for publication bias (Sterne, Egger, & Moher, 2011). In addition, we explored Egger's test of asymmetry.

Results

Study Selection

Study selection results are displayed in Figure 2. Through searches of electronic databases and grey literature (e.g., theses, dissertations and conference papers) 6,456 records were identified. After reviewing the titles and abstracts of these 6,102 non-duplicate records, 757 potentially relevant full-text records were obtained and reviewed. After full-text review, 176 studies met inclusion criteria and were included in the meta-analysis. Raw data can be downloaded from: https://osf.io/mcavt/?view_only=dbd862b10bb2412fa908251198e5d18f

Study Characteristics

Study characteristics are detailed in supplementary materials. Publication dates ranged from 1966 to 2016. Most of samples came from the last 20 years (n [number of study samples]: <1980s = 15; 1908s = 19; 1990s = 65; 2000s = 80; \geq 2010 = 76). Participant mean age ranged from 7.04 years to 33.30 years. Samples were categorized according to age group: elementary school age ($n = 33$), middle school age ($n = 73$), high school age ($n = 85$), young adult ($n = 54$) and adult ($n = 6$).

Most studies were conducted in the United States; however, the review includes studies from Asia, Africa, South America, Oceania and Europe. Countries were classified according to the United Nation's Gender Inequality Index 2014 (GII). Most samples were from countries with either very high gender equality ($n = 27$), high gender equality ($n = 60$) or medium ($n = 148$) level of gender equality, with a smaller number of samples from countries with low ratings of gender equality ($n = 13$).

During data extraction, information was collected about the social class of the sample. Most study samples were described as either lower SES (e.g., majority working class or lower-middle-class, $n = 35$) and middle-class (n

= 28). There were only nine samples based on high SES samples. The percentage of participants identifying as belonging to an ethnic minority was skewed towards samples that were lower in ethnic diversity. There were 58 samples from samples with less than 25% of participants belonging to an ethnic minority, however, 17 effect sizes came from 25-49% ethnic minority samples, and 36 samples were from samples with a majority of students identifying as belonging to an ethnic minority.

Missing Study Categories

Hyde (2013) has claimed that many studies do not include sufficient information on demographic characteristics to even explore intersectional perspectives. We found strong support for this. A remarkable feature of in the current review is that we could only extract SES data from 28.1% of study samples and ethnic minority status of 43.4% of study samples. No such issue was found for age group where we had complete information. This signifies the need for better reporting of demographic characteristics.

Synthesis of Results

Math expectancy for success and value. In support of Hypothesis 1, boys had higher levels of expectancy and value in math (see Table 1 for full results of pooled effect sizes across all domains). The overall pooled effect of gender on math expectancy for success was $d = 0.27$, 95% CI [0.23, 0.31], indicating that boys had a small advantage in terms of their perceived ability to do well in math. Effects for the different components of math task value were smaller but were still in the hypothesized direction. Math task value ($d = 0.14$, 95% CI [0.06, 0.21]), and math intrinsic value ($d = 0.17$, 95% CI [0.12, 0.22]), favored boys the most out of all the task value components. Whereas, gender differences for math utility value ($d = 0.08$, 95% CI [0.02, 0.13]), math attainment value ($d = 0.02$, 95% CI [-0.10, 0.15]), and math cost ($d = -0.08$, 95% CI [-0.21, 0.05]), were negligible. Generally, there was a substantial degree of heterogeneity across the effects for gender expectancy/value (I^2 ranging from 0.84-0.93). The only exception to this was for math cost². Figure 2 presents an example funnel plot. Funnel plots for all other variables can be found in supplementary materials.

Science expectancy for success and value. Gender differences were weaker for general science expectancy/value than they were in more precisely defined areas of STEM (e.g., general science versus physical or biological sciences). The overall pooled effect of gender on science expectancy for success was $d = 0.18$, 95% CI [0.10, 0.26], indicating that boys had a slight advantage in terms of their perceived ability to do well in science. The strongest effect was science intrinsic value $d = 0.21$, 95% CI [0.11, 0.32], whereby boys were more likely to report

higher intrinsic value in science. There was little or no difference between genders in science task value ($d = 0.01$, 95% CI [-0.06, 0.08]), science utility value ($d = 0.05$, 95% CI [-0.02, 0.12]), and science attainment value ($d = 0.05$, 95% CI [-0.02, 0.13]). Overall, there was a large amount of heterogeneity across effects with I^2 ranging from 0.60 - 0.88. There were not enough studies to provide a meta-analysis on science cost, but available effect sizes indicated that there was a small effect of girls being more likely to report higher levels of science cost.

Computing expectancy for success and value. In support of Hypothesis 1, effects of gender on computing expectancy for success and value were in favor of boys. The overall pooled effect of gender on computing expectancy for success was $d = 0.44$, 95% CI [0.28, 0.60], demonstrating a medium sized effect favoring boys in terms of their perceived ability to do well in computing. In line with the aforementioned results, intrinsic value showed larger gender effects in comparison to other components of task value ($d = 0.48$), 95% CI [0.26, 0.69]. Effect sizes for computing task value ($d = 0.22$), 95% CI [0.07, 0.38], and computing utility value ($d = 0.21$), 95% CI [0.07, 0.35] also revealed a small effect for girls being less likely to rate computing as high in task value and career/practical values in comparison to boys. There was a large degree of variance across effect sizes with I^2 ranging from 0.86 - 0.97.

Engineering expectancy for success and value. Again, in support of Hypothesis 1, effects of gender on engineering expectancy for success and value were in favor of boys. The overall pooled effect of gender on computing expectancy for success was $d = 0.24$, 95% CI [0.08, 0.40], demonstrating a small effect favoring boys in terms of their perceived ability to do well in engineering. Engineering intrinsic value showed a similar effect ($d = 0.22$, 95% CI [0.11, 0.32]), however the effect of gender on engineering utility value was negligible ($d = 0.04$, 95% CI [-0.14, 0.22]). I^2 scores varied considerably, most likely due at least in part to the small number of studies on engineering expectancy and value.

Physical sciences expectancy for success and value. The overall pooled effect of gender on expectancy for success in the physical sciences was $d = 0.43$, 95% CI [0.29, 0.56], demonstrating a medium effect favoring boys in terms of their perceived ability to do well in the physical sciences. Again, out of all the task value components intrinsic value showed the largest effect: ($d = 0.27$, 95% CI [0.19, 0.36]). In keeping with the above results, the effect for utility value was extremely small ($d = 0.05$, 95% CI [-0.11, 0.2]). In contrast to predictions, physical science task value showed a small, but non-significant effect in favor of girls ($d = -0.14$ 95% CI [-0.66, 0.38]). Physical science cost was showed a medium effect for gender, with girls reporting higher levels of cost for engaging

with physical sciences ($d = -0.32$, 95% CI [-0.43, -0.21]). I^2 scores varied considerably, but values were interpreted with caution due to the low number of studies including physical science (see Table 1 for number of effect sizes).

Biological sciences expectancy for success and value. In contrast to hypotheses, there was almost no effect of gender on expectancy for success in the biological sciences ($d = -0.03$, 95% CI [-0.19, 0.10]). However, in support of hypotheses there was a positive effect of being a girl in terms of interest in biological sciences ($d = -0.23$, 95% CI [-0.40, -0.06]). Again, there was a smaller non-significant effect for utility value ($d = -0.09$, 95% CI [-0.30, 0.12]). I^2 scores ranged from 0.60 to 0.90 indicating considerable heterogeneity across the effect sizes. It is worth noting that biological science did have a significant Egger's test (2.77, $p = .015$). This may, however, be due to the few effects sizes available for analysis for this academic discipline.

Verbal expectancy for success and value. In support of hypothesis I, effects of gender on verbal expectancy for success and value were in favor of girls. The overall pooled effect of gender on verbal expectancy for success was $d = -0.17$, 95% CI [-0.23, -0.11], demonstrating a small effect favoring girls. Verbal task value showed the largest effect ($d = -0.48$, 95% CI [-0.62, -0.34]), followed by verbal intrinsic value ($d = -0.32$, 95% CI [-0.40, -0.24]), In comparison to task value and intrinsic value, utility ($d = -0.27$, 95% CI [-0.31, -0.23]) and attainment value ($d = -0.28$, 95% CI [-0.35, -0.22]) had comparatively lower effects. I^2 scores varied considerably for verbal expectancies for success and value, however, values were likely affected by the low number of studies included in analyses.

Substantive Moderators – Sample Level

Moderator analyses for social class and ethnicity should be interpreted by the reader with some caution given the missing data on these variables that characterizes this literature. We focus our attention here only on significant moderation for each moderator given the number of analyses undertaken (see Table 2). All moderation results can be found in Table S1 in supplementary material. The coding scheme for the moderator variables can be found in supplementary material.

Social class. Gender gaps were largest in high SES samples, and lowest in low SES samples for nearly all math variables. For instance, for math expectancy the gender gap rose from a small effect for majority working class samples: ($d = 0.19$, 95% CI [0.28, 0.10]), to a slightly larger effect for majority middle-class samples ($d = 0.25$, 95% CI [0.33, 0.17]), and finally, to a large effect for the most affluent samples ($d = 0.77$, 95% CI [1.24, 0.31]).

Importantly, although confidence intervals in each category overlapped, the moderation effect for social class was

statistically significant at $p = 0.00$. This effect was similar for math task value ($p = 0.00$); majority working class samples: ($d = 0.03$, 95% CI [0.12, -0.06]), compared to majority high SES samples: ($d = 0.67$, 95% CI [1.00, 0.34]); and also math intrinsic value; majority working class samples: ($d = 0.05$, 95% CI [0.11, -0.01]) and majority middle-class samples: ($d = 0.08$, 95% CI [0.20, -0.04]); compared to the majority high SES samples: ($d = 0.47$, 95% CI [0.75, 0.19]). Math utility also showed a trend towards the same direction; majority working class samples: ($d = -0.05$, 95% CI [0.04, -0.14]), compared to majority middle-class samples: ($d = 0.10$, 95% CI [0.20, 0.01]). Less data was available to analyze the effect of social class on other domains. Science expectancy for success showed a similar statistically significant moderation effect for social class; majority working class samples: ($d = 0.14$, 95% CI [0.27, 0.01]), compared to majority middle-class samples: ($d = 0.56$, 95% CI [0.86, 0.27]). Finally, there was no significant interaction in verbal domains. However, it should be noted that there was substantially less data available to test effects of social class.

Ethnic minority percentage of sample. Overall, there were few effects that were moderated by the proportion of participants who identified as belonging to an ethnic minority. The only exception was for math utility value. For math utility value, samples with the lowest ethnic diversity had the largest gender gaps favoring boys: ($d = 0.13$, 95% CI [0.23, 0.03]). In contrast, samples with high ethnic diversity had gender gaps that favored girls ($d = -0.16$, 95% CI [-0.07, -0.24]).

Age. Age was generally not a significant moderator of gender effects. This may be because there were relatively few studies on young children or adult student populations. Rather, most studies focused on middle and high-school aged children. The one significant interaction was for science expectancy for success ($p = 0.03$), where the gender gap was larger for older samples. In this instance, the pooled effect for gender on science expectancy for success was almost non-existent for samples comprised of elementary school children: ($d = -0.05$, 95% CI [0.03, -0.14]), however, rose to ($d = 0.31$, 95% CI [0.51, 0.11]) for young adult samples.

Sample type. There was little evidence for any moderation effect across sample type (e.g., normal samples versus elective, university or gifted students). The one exception was for computing expectancy for success, whereby students in elective or advanced courses showed larger gender differences ($d = 0.61$, 95% CI [0.83, 0.39]), versus students from 'normal' populations ($d = 0.23$, 95% CI [0.41, 0.05]).

Country Level Indicators of Gender Equality. We aimed to replicate the country-level moderator analysis of Else-Quest et al. (2010). Here we included both the composite measures and the domain specific

measures used by Else-Quest and colleagues to compare both the gender stratification hypothesis (i.e., greater gender equality will lead to lower STEM gaps) and the gender equality paradox (i.e., high gender equality will lead to greater gender gaps). Unlike the other moderators we rely on summaries of the results rather than providing interpretation of individual significant effects (which can all be found in Table 3; complete findings can be found in supplementary materials Tables S2). This is because there were 95 separate moderation models. Of these 95 models only 27 (28%) were statistically significant. Of these 13 were for the composite measures and 14 were for the domain specific measures. In both cases the average absolute effect size was modest (.07 and .09 for composite and domain specific measures respectively). Domain specific measures ranged from -.11 for the moderation of science utility value gender differences by the tertiary ratio of female to male enrolment ratio to .34 for the moderation of physical science expectancy gender differences in labor force participation. Composite measures also had effect sizes that ranged from negative (-.14 for computing expectancy gender differences by the GGI) to positive (.11 for physical science expectancy gender differences by the GGI). Generally, composite measures tended to favor the gender equality paradox, while domain specific measures were more mixed with four favoring the gender stratification hypothesis and the rest supporting the gender paradox. This patchwork of significant and insignificant findings and variation in direction of effects is quite similar to the pattern found in Tables 6 and 7 of Else-Quest and colleagues.

Publication date (era). For nearly all variables, there was little variation in effect sizes according to the era in which the studies were published, with the exception of math utility and verbal intrinsic value. For math utility value, the gender difference that favored boys in the 1980s ($d = 0.18$, 95% CI [0.30, 0.06]) all but disappeared from the 2000s onwards ($d = -0.03$, 95% CI [0.05, -0.10]) for the 2000s, and ($d = 0.02$, 95% CI [0.15, -0.10]) for 2010 and beyond. Gender differences favoring girls in verbal intrinsic value in the 1990s were larger ($d = -0.40$, 95% CI [0.30, -0.51]), compared to results from 2010 onwards ($d = -0.21$, 95% CI [-0.08, -0.34]). Limited data for earlier time periods prior to 2000s meant that it was only possible to test the effect of publication date for a subset of variables.

Reliability. There was little evidence for gender differences varying across psychometric ratings of reliability. Math expectancy for success and verbal intrinsic value were the only variables that had a significant moderation effect for reliability, however these effects went in opposite directions (e.g., math expectancy for success showed that studies with low reliability were more likely to report smaller gender differences, whilst verbal intrinsic

value showed the opposite). Taken together there was little evidence that psychometric ratings of reliability moderated the size of gender differences.

Discussion

This study used a meta-analytic approach to examine gender differences and similarities across a wide range of EVT constructs in a variety of different domains. Additionally, the meta-analysis provided an opportunity to synthesize current literature through an intersectional lens, by exploring the degree of heterogeneity in effect sizes and the degree to which gender differences varied across a number of cultural and social contexts. These findings shed light on a number of issues pertinent to the study of gender and attitudes towards math, science, and verbal domains. As expected, gender differences followed a gender stereotypical pattern across academic domains. Yet our evidence suggests that heterogeneity in effects sizes was large and there were important moderators that help explain some but by no means all of this heterogeneity.

Gender Differences and Similarities Across EVT Constructs

Seventeen out of the 31 pooled effect sizes in this study were small ($d \sim .20$). Four effects were medium in size ($d \sim .50$). None were large. Hyde's (2005) gender similarity hypothesis states that gender differences in many, but not all, variables will be small and practically non-significant. For academic constructs, differences tend to be small. This was certainly the case here. The absolute average effect size in our study was at the threshold between a small and trivial effect sizes ($|M| = .200$). No effect size was larger than .50 and approximately 30% of the pooled effect sizes we explored were not significant. An intersectional perspective also requires us to consider not only the average effect size but also the degree of heterogeneity in effect sizes. In this case, for almost all effect sizes the heterogeneity was large. Although we discuss moderation below, it is worth noting that heterogeneity often remained large within the moderator categories. It is this combination of small effect sizes and large heterogeneity that makes Hyde's gender similarity hypothesis such a compelling account of the findings in the literature. Yet, the sheer number of effects in this meta-analysis—remembering that we considered only those studies that had both an expectancy and a value variable—show that gender differences in EVT variables remain a hot topic.

A critical realist approach to intersectionality suggests that studying gender is useful because, although gender does not provide a compelling account of any particular individual, it does provide insight into a structural position (Gunnarsson, 2011). However, at least in the case of EVT variables, gender may not be as powerful an abstraction as the amount of literature would lead us to think. As such, we suggest that a) research should now pursue an intersectional approach to gender differences in EVT (there is perhaps enough research that considered

gender differences, unconnected from other contextual factor or social identity, in EVT variables) and b) that greater focus may need to be given to other factors like ethnicity and social class that seek to uncover other sorts of power relationships (see for example Parker et al., 2018). If gender differences in EVT variables appear to be small and heterogeneous then why do differences in female undergraduate enrolments remain in some STEM fields? Cheryan, Ziegler, Montoya, and Jiang's (2016) suggest that masculine cultures in some STEM fields and a lack of experience with computer science, engineering, and physics for girls provide powerful explanations for such underrepresentation.

Domain and Dimension Specificity

A key finding was that effect sizes for gender differences in academic attitudes were highly domain specific (e.g., computer science vs biology) and that effect sizes varied within domain by EVT dimension (e.g., expectancies vs value). When considering domain, gender differences in general science were generally small. In contrast, gender differences in the physical sciences and computing were comparatively larger favoring boys. Small differences favoring girls were evident in the biological sciences as were larger differences favoring girls in verbal domains. Cheryan, et al.'s (2016) review argues that research needs to focus not just on women's trajectories but on the specific domains of STEM. Largely, women's STEM gap is dependent on how STEM is defined. Indeed, the STEM gap may reflect perceptions that physics, engineering, and computer science—fields where men tend to be relatively well represented—are considered “real” sciences, while biology, medicine, and indeed the social sciences—fields where women are better represented—are not. Our results suggest researchers should be clearer in what domains they are talking about when discussing the STEM gender gap. Indeed, it may well be time to retire the phrase STEM gender gap as it so clearly fails to account for domain to domain nuances. In addition, our results suggest that general science variables are much less likely to reveal a nuanced understanding of gender differences when compared to field specific measures. Where appropriate researchers should use domain specific measures.

An additional finding was that gender differences within domains varied widely by EVT dimension. Gender differences appeared to be larger for self-evaluative (e.g., expectancies) and affective (e.g., intrinsic value) variables. In contrast, differences were smaller for more extrinsic variables like utility value and cost. This suggests that girls and boys do not differ very much in their understanding of why STEM is important for future attainment or what costs are involved. Rather differences tend to emerge in boys' and girls' intrinsic relationship with STEM fields. Thus, to the degree that EVT variables contribute to gender differences in STEM attainment it is likely not

due to differences in rational assessments of costs and benefits. The average effect size, even for intrinsic variables, was typically small and heterogeneity large. Yet larger effect sizes for expectancy and intrinsic value were observed for computing and physics. Above we suggested that gender may not be a particularly powerful social category to consider for EVT variables in general. However, when considering both dimension and domains together we suggest that gender differences in computing and physics—and potentially engineering—expectancies and intrinsic value remains an important research topic and may help explain lower attainment in these fields (Cheryan et al., 2016).

Moderation Effects for Social and Cultural Contexts

Gender and social class. The lack of information reported on sample characteristics for social class or ethnicity we witnessed in our review of the data is problematic for several reasons. First, it makes meta-analyses aimed at considering the effects of multiple aspects of social identity difficult. Second, it suggests that researchers are not attuned to the importance of multiple categories of social identity. Where data was available the most interesting findings were found in relation to social class. We emphasize here that we use the term social class loosely as we were dependent on study authors' description of their sample. Sample descriptions tended to use thin descriptions of social class, did not distinguish between economic, cultural, or social capital, nor did they distinguish social class from socioeconomic status.

Nevertheless, social class showed a number of significant moderator effects across math related variables and science expectancy for success, whereby the gender gap favoring boys increased as social class became higher. This finding reflected the results of Fryer and Levitt (2010), Lubienski, Crane, and Robinson (2011) and McGraw et al. (2006) who found that the gender achievement gap in math for elementary and high school students is larger amongst high SES students compared to low SES students. As we will discuss below there is considerable tension in the literature about the effect of macro-country level contextual effects on STEM related variables. Yet, there is a need to consider within country contextual effects. As intersectionality would suggest, individual's contexts are not merely reflected by nationality but are bisected by ethnicity, social-class, and a number of other social identities.

Why is there such a large gender gap for people in higher socio-economic statuses versus those from majority working class populations? Due to the inherent weaknesses of a meta-analysis approach to moderation analyses such as limited moderator data at the individual level (see Author, 2009 for a discussion), results should be interpreted with caution. Nonetheless, primary research directly assessing this issue found evidence that socio-economic status moderates the size of gender differences in math self-efficacy—even when controlling for academic

achievement (Van Zanden, 2018). However, this interaction did not extend to several other math self-beliefs and attitudes.

A potential hypothesis behind the gender/social class relationship could be that children from wealthier, resource-rich families are provided with more gender socialization opportunities. Indeed, there is evidence to suggest that children from higher socio-economic statuses are exposed to more gender specific parenting patterns and extra-curricular activities outside of school, compared to their peers from lower SES families (Lareau, 2003). An unanticipated negative effect of this is that exposure to gender stereotypical activities results in greater gender stereotypic differentiation in the self-beliefs and attitudes of children. As children gain more experience in gender-congruent activities, and less experience and familiarity with tasks and activities incongruent with their gender, their self-beliefs and attitudes consequently become confined to a gender stereotypical pattern reflecting their exposure (or lack of) to different experiences. Interestingly, this pattern occurs despite the fact that parents from high SES are more likely to claim to hold gender egalitarian views (Marks, Lam, & McHale, 2009).

Another alternative explanation is that high SES environments often have higher levels of average achievement and this can translate into a more competitive environment for students. A growing body of research has shown that on average women respond less favorably to competitive environments than men (e.g., Bönke 2015; Gneezy, Niederle, & Rustichini, 2003; Niederle & Vesterlund 2007, 2010). In the context of education, Alon and DiPrete (2015) showed that the intensity of competition, as signaled by admission standards into university STEM courses, had a larger deterring effect on girls than boys.

It is, however, intriguing that the same pattern did not extend to all domains. One reason for this is that there was limited data available to test social class moderation effects in all cases. Thus, research should endeavor to include social class in participant demographics, and to recruit a wider range of participants outside of the middle-class demographics that has traditionally been oversampled in psychological research.

Gender and ethnicity. We explored the relationship between gender and ethnicity by comparing the size of gender differences across samples with differing levels of ethnic diversity. Importantly, there were more similarities than differences in effect sizes amongst samples with students from a range of different ethnic backgrounds and samples with little ethnic diversity, with the exception of some differences showing that participants from samples with low ethnic diversity were more likely to have gender differences favoring boys in math utility value. These low ethnic diversity samples were primarily white. Overall, results reflect previous

research that has provided mixed evidence regarding the interaction between gender and ethnicity. While some research has suggested that gender differences in attitudes and achievement become larger amongst white students (e.g., Catsambis, 1994; 1995; Cooley, 2001; NSF, Division of Science Resource Statistics, 2012; McGraw et al., 2006; Zarrett et al., 2006), other studies have not replicated this effect (e.g., Else-Quest et al., 2013; Zarrett et al., 2006). Importantly, these results are limited by the fact that many different ethnicities may exist within the category of ethnic minority, and therefore, important differences may be overlooked by the use of such a broad category.

Moderation Effects for Country Level Indicators

There are two main competing theories of the effect of country context on gender differences in STEM variables. The gender stratification hypothesis argues that gender differences in STEM achievement variables are due to differences in the political, economic, or cultural context that structure individual's choice architecture and that as gender equality increases women are afforded greater freedom to consider STEM as a viable option and come to view their abilities in this area more positively (Else-Quest et al., 2010). The gender equality paradox, however, argues that empirical research actually finds the opposite effect (Stoet & Geary, 2018). Before discussing our results, it is worth noting several weaknesses in the existing research. First, the gender equality paradox appears to be largely supported by research that uses composite metrics of country level gender inequality. These measures have questionable theoretical and psychometric underpinnings (Hawkin & Munck, 2013). In contrast, research on the gender stratification hypothesis (e.g., Else-Quest et al., 2010) has used a greater variety of measures, including more concrete domain-specific measures, but the results using these measures has resulted in a patchwork of significant and non-significant findings that appear to have no clear pattern. Using a similar list composite and domain specific of indicators as Else-Quest et al. (2010) our results are similar. Composite measures tended to favor the gender equality paradox, while domain specific measures tended to have more mixed findings. In total, however the effect sizes were small and the vast majority of moderation effects—almost three-quarters—were not significant. Taken together, we suggest that the inconsistency in effects suggests that there is no clear relationship between national gender equality and STEM gender differences (see also Ireson, 2017). This may be due to domain specific measures of the STEM context being unable to fully capture the full cultural milieu that influences boys and girls STEM beliefs and the questionable construct validity of broader composite measures of the gender equality context (Hawkin & Munck, 2013). Better measurement work in this area is needed.

Moderation Effects for Study Characteristics

Age. Another key finding was that the effect of gender was unrelated to average participant age, with the exception of science expectancy for success. Again, this finding is surprising, given that the age range of participants in the meta-analysis ranged from 7 to 33 years. One conclusion of this finding could be that gender socialization has a limited role in determining the degree to which self-beliefs and attitudes are differentiated according to gender. Indeed, the gender stratification hypothesis maintains that gender differences should become larger across development because of greater exposure to gender socialization. An alternative to this explanation could be that gender socialization is so heavily entrenched during early childhood that gender differences emerge at ages younger than what this study examined. Regardless, these findings show that gender stereotypical patterns in self-beliefs and attitudes are heavily entrenched even in early childhood. One gap in the literature is that there is an over-representation of school age children, but limited attention to children below school age. Given that gender differences are already established amongst elementary aged children, researchers need to start to focus on the emergence of gender differences in self-beliefs and attitudes in children before they engage in formal schooling. While there are likely many methodological challenges in working with such a young population, this work is critical to furthering our understanding of gender differences in self-beliefs and attitudes.

Publication date (era). Another way of exploring the role of cultural context in relation to gender differences in attitudes and beliefs is assessing the relationship between publication date and effect sizes. For example, Parker et al. (2018) showed that gender gaps in self-concept have historically increased as achievement gaps have tended to decline. Yet in our research, there were few statistically significant effects by publication date. The only exception was for math utility value, whereby effect sizes have diminished since the 1980s to almost zero difference in the current day. A similar pattern emerged for verbal intrinsic value, whereby gender differences favoring girls decreased somewhat from the 1990s to the current day. The lack of change in gender differences in attitudes for other variables paints a potentially damning picture of gender equality in educational attitudes, but we are cautious to over-interpret these findings. Firstly, there was limited data available for the time period before the 1980s, with most studies being published within the last 15-20 years. Thus, this meta-analysis was unable to provide conclusive evidence on whether the gender gap in self beliefs and attitudes has changed since much earlier decades in the twentieth century. Nonetheless, it is concerning that there has been little change since the 1980s for most variables. Indeed, this is in line with recent research that has shown the persistence of gender stereotypes across time, showing that people's perceptions of gender stereotypes have been largely stagnant and resistant to change

since the 1980s (e.g., Haines, Deaux, & Lofaro, 2016). However, the results are still more promising than those of Parker et al. (2018) which show historical trends upward in gender difference. This may, however, be because Parker et al. considered equally able boys and girls while our meta-analysis had no such achievement controls.

Sample type. Moderation effects across sample types also revealed some critical insights into contextual factors and gender differences. Analyses showed that for nearly all variables there were no significant difference between ‘normal’ samples of students versus samples drawn from students taking advanced or elective courses. Nonetheless, there was one significant effect for engineering expectancy for success that showed that the gender effect (favoring boys) was largest in elective/advanced samples. The lack of moderation by sample type suggests that even for girls who study STEM studies at a high level face challenges in terms of lower self-beliefs and task values. Thus, educators should be aware that even amongst high-achieving and highly motivated populations, girls still have lower confidence and poorer attitudes towards STEM relative to boys. Again, we wish to point the reviewer to the small effect sizes and large heterogeneity.

Boys and verbal domains. Finally, this meta-analysis also highlights the disadvantage that boys face with respect to verbal domains. Indeed, effects sizes were generally larger here than in the STEM fields—though still modest in size and still heterogeneous. Thus, educational policies that discuss gender equality in education need to also recognize this. If we are to encourage truly gender equal education, we need to view gender equality holistically, considering both verbal and STEM domains when discussing differences in self-beliefs and attitudes so all children develop perceptions of their abilities commensurate with their achievement. The general weak effect sizes across the board however suggest that policy may well be better served by considering early experiences and masculine/feminine culture of different academic domains as more promising areas of intervention (Cheryan et al., 2016).

Limitations and Directions for Future Research

This study provided gender differences and similarities across a diverse range of expectancy value constructs in a number of domains. This study also explored the use of meta-analytic techniques within the framework of an intersectional perspective to gender differences. By gathering data about the social class, ethnicity, and country of origin for each study, more insight was gained into how social forces such as class and different aspects of gender equality are associated with gender differences in self-beliefs and attitudes across math, science, and verbal domains. Results showed a relationship between social class and gender that consistently appeared across

math attitudes, showing that more affluent girls are potentially prone to greater disadvantage in academic self-beliefs and attitudes relative to similar positioned boys. There was a complex relationship between gender equality, attitudes, and gender segregation in educational outcomes by showing that the gender ratio of girls enrolled in STEM versus humanities was a stronger predictor of gender differences, than the broader GII metric commonly used by other researchers. However, there were a number of difficulties that place some limitation on the conclusions that can be drawn from the study, particularly with regards to using meta-analytic methods to explore intersectional questions.

Firstly, this review has revealed that many studies did not include basic demographic information about their samples, particularly in relation to social class and ethnicity. This lack of information made it difficult to gain enough power to perform moderation analyses on all variables included in the meta-analysis. As a result, the ethnicity analysis had to be restricted to the percentage of participants as belonging to an ethnic minority, as opposed to particular ethnic groups. Results indicated that there is likely to be more similarities than differences amongst ethnic groups for gender differences, however, this conclusion is limited in that amongst any given ethnic minority there is a wide variation of cultures and practices, that make the category of ethnic minority extremely heterogeneous. Therefore, it is difficult to ascribe reliable conclusions from the data on ethnicity. Furthermore, categorization of gender equality was also limited in that it was difficult to match GII ratings specifically to the year of publication because of power needed to perform moderation analyses and the GII's relatively recent history. Nonetheless, gender inequality of nations (and other country-level indicators), is unlikely to dramatically change across the years (Stotsky, Shibuya, Kolovich, & Kebhaj, 2016). Thus, results from this study are tentative in that they are based on less precise information compared to other research using data that matches directly to the year in which the gender indices were reported.

This review also revealed that there is a substantial over-representation of white American middle-class participants in the literature, confirming wider critiques of psychological research (e.g., Henrich et al., 2010; Hyde, 2007). As such, most of the literature on expectancy value theory and gender is actually based upon a homogenous group of people. Future research should be mindful of the need to extend research agendas to sample beyond such narrow demographics, and to promote further research that is inclusive of a diverse range of populations and experiences. Additionally, researchers should aim to report more detailed demographic information relating to the ethnicity and socioeconomic status of participants.

Finally, this review highlighted gaps within the expectancy value literature. In comparison to other subjects, math was over-researched. The domain specific aspects of STEM (e.g., engineering, physical sciences, biological sciences, and computing) were under-researched in comparison to general measures of math and science. Given that gender differences are likely to be better understood using domain specific STEM measures, it is critical that researchers researching the gender STEM gap consider moving away from general science and math measures, in favor of exploring domain specific measures (e.g., physics self-concept). In addition to this, more research focusing on boys' verbal expectancies for success and task values needs to be done.

The final research gap that was highlighted was the components of expectancy value that were included in the study. Some aspects of expectancy value were under-researched, leading to variables such as cost, and attainment value being neglected in the study of gender differences in education. Further research should explore these constructs in greater depth to better understand student motivation.

Conclusion

In summary, this review has demonstrated the importance of exploring gender differences in relation to social and cultural context, as well as assessing differences across sub-disciplines within STEM. Hopefully this research encourages more studies to begin investigating gender in relation to other social categories, and to further explore and develop new research methods that address the limitations of a quantitative approach to intersectionality. Understanding the social and cultural contexts in which gender differences in self-beliefs and attitudes vary will ultimately help educators identify the particular groups of students who would benefit the most from interventions to decrease gender disparities in educational outcomes. Furthermore, by better understanding the role of social and cultural contexts, we can come a step closer to understanding the environments in which gender gaps in education become most problematic, and those in which gender gaps are at their smallest, thus, paving the way for future research that identifies the most beneficial environments for all students.

Footnotes

¹Heterogeneity estimates for Math Cost were regarded with caution because of the small number of studies in analyses (see von Hippel, 2015 for a discussion of I^2 biases in small meta-analyses).

² Heterogeneity estimates for Math Cost were regarded with caution because of the small number of studies in analyses (see von Hippel, 2015 for a discussion of I^2 biases in small meta-analyses).

References

References in the Meta-analysis are in Supplementary Materials

- Alon, S., & DiPrete, T. A. (2015). Gender differences in the formation of field of study choice set. *Sociological Science*, 2, 50-81. doi:10.15195/v2.a5
- Auwarter, A. E., & Aruguete, M. S. (2008a). Effects of student gender and socioeconomic status on teacher perceptions. *The Journal of Educational Research*, 101, 242-246. doi:10.3200/joer.101.4.243-246
- Baker D. P., Jones D. P. (1993) Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education*, 66, 91–103. doi:10.2307/2112795
- Bönte, W. (2015). Gender differences in competitive preferences: New cross-country empirical evidence. *Applied Economics Letters*, 22, 71-75. doi:10.1080/13504851.2014.927560
- Bowleg, L. (2008). When Black+ lesbian+ woman ≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex roles*, 59, 312-325. doi: 10.1007/s11199-008-9400-z
- Catsambis, S. (1994). The path to math: Gender and racial-ethnic differences in mathematics participation from middle school to high school. *Sociology of Education*, 199-215. doi:10.2307/2112791
- Catsambis, S. (1995). Gender, race, ethnicity, and science education in the middle grades. *Journal of Research in Science Teaching*, 32, 243-257. doi:10.1002/tea.3660320305
- Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*. 114, 924-976. doi:10.1086/595942
- Charles, M., Harr, B., Cech, E., & Hendley, A. (2014). Who likes math where? Gender differences in eighth-graders' attitudes around the world. *International Studies in Sociology of Education*, 24, 85-112. doi:10.1080/09620214.2014.895140
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others?. *Psychological Bulletin*, 143, 1-35. doi: 10.1037/bul0000052
- Cheung, M. (2011). *metaSEM: meta-analysis: A structural equation modelling approach*. R package version 0.5-3. Retrieved from <http://courses.nus.edu.sg/course/psycwlm/Internet/metaSEM>

- Cheung, M. (2014). Modeling dependent effect sizes with three-level metaanalyses: A structural equation modeling approach. *Psychological Methods, 19*, 211-229.
doi:10.1037/a0032968
- Choo, H. Y., & Ferree, M. M. (2010). Practicing intersectionality in sociological research: A critical analysis of inclusions, interactions, and institutions in the study of inequalities. *Sociological Theory, 28*, 129–149.
doi:10.1111/j.1467- 9558.2010.01370.x
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coley, R. J. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Policy Information Report.
- Collins, P.H. (1990). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. New York: Routledge.
- Collier, A. (1994). *Critical realism: An introduction to Roy Bhaskar's philosophy*. London: Verso.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *U. Chi. Legal F., 139*.
- Eagly, A. H., Eaton, A., Rose, S. M., Riger, S., & McHugh, M. C. (2012). Feminism and psychology: analysis of a half-century of research on women and gender. *American Psychologist, 67*(3), 211. doi:10.1037/a0027260
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly, 18*(4), 585-609. doi:10.1111/j.1471-6402.1994.tb01049.x
- Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. *Handbook of competence and motivation*, 105-121.
- Eccles, J. S., & Hoffman, L. W. (1984). Socialization and the maintenance of a sex-segregated labor market. In H. W. Stevenson & A. E. Siegel (Eds.), *Research in Child Development and Social Policy* (Vol. 1, pp. 367-420). Chicago: University of Chicago Press.
- Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11*, 367-380.
doi:10.1086/494229
- Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*, 183-201. doi:10.1111/j.1540-4560.1990.tb01929.x

- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132. doi:0.1146/annurev.psych.53.100901.135153
- Else-Quest, N. M., & Grabe, S. (2012). The political is personal: Measurement and application of nation-level indicators of gender equity in psychological research. *Psychology of Women Quarterly*, *36*, 131-144. doi:10.1177/0361684312441592
- Else-Quest, N.M. & Hamilton, V. (2018) Measurement and analysis of nation-level gender equity in the psychology of women. In C.B. Travis & J.W. White (Eds.) *APA handbook of the psychology of women (pp. 545-563)*. Washington, DC: APA Press
- Else-Quest, N. M., & Hyde, J. S. (2016a). Intersectionality in quantitative psychological research I. Theoretical and epistemological issues. *Psychology of Women Quarterly*, *40*, 155-170. doi:10.1177/0361684316629797
- Else-Quest, N. M., & Hyde, J. S. (2016b). Intersectionality in quantitative psychological research II. Methods and techniques. *Psychology of Women Quarterly*, *40*, 319-336. doi:10.1177/0361684316647953
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*, 103. doi:10.1037/a0018053
- Few-Demo, A. L. (2014). Intersectionality as the “new” critical approach in feminist family studies: Evolving racial/ethnic feminisms and critical race theories. *Journal of Family Theory & Review*, *6*, 169–183. doi:10.1111/jftr.12039
- Field, A. P. (2003). The problems in using fixed-effects models of metaanalysis on real-world data. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *2*, 105-124. doi:10.1207/S15328031US0202_02
- Fox, D., Prilleltensky, I., & Austin, S. (Eds.). (2009). *Critical psychology: An introduction*. London, UK: Sage.
- Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*, 109–118. doi:10.1037/0003-066X.56.2.109
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, *118*, 1049-1074. doi:10.1162/00335530360698496
- Goldstein, H. (1995). *Multilevel statistical models*. London, UK: Edward Arnold.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*, 1164-1165. doi:10.1126/science.1154094

- Gunnarsson, L. (2011). A defence of the category 'women'. *Feminist theory*, *12*, 23-37. doi: 10.1177/1464700110390604
- Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., & Van Zanden, B. (2019). Countries, parental occupation, and girls' interest in science. *The Lancet*, *393*(10171), e6-e8. doi: 10.1016/S0140-6736(19)30210-7
- Hancock, A. (2016). *Intersectionality: An intellectual history*. Oxford University Press: Kindle Edition.
- Haines, E., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing ... or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*, 353–363, doi:10.1177/0361684316634081
- Hawken, A., & Munck, G. L. (2013). Cross-national indices with gender-differentiated data: what do they measure? How valid are they?. *Social indicators research*, *111*, 801-838. doi: 10.1007/s11205-012-0035-7
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi:10.1017/ S0140525X0999152X
- Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560. doi:10.1136/bmj.327.7414.557
- Hull, G. T., Scott, P. B., & Smith, B. (1982). *All the women are white, all the men are black, but some of us are brave*. New York: Feminist.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581. doi:10.1037/0003-066X.60.6.581
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, *16*, 259-263. doi:10.1111/j.1467-8721.2007.00516.x
- Hyde, J. S. (2012). Nation-level indicators of gender equality in psychological research: Theoretical and methodological issues. *Psychology of Women Quarterly*, *36*, 145-148. doi:10.1177/0361684312441448
- Hyde, J. S. (2013). Gender similarities and differences. *Annual review of psychology*, *65*, 373-398. doi:10.1146/annurev-psych-010213-115057
- Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, *106*, 8801-8807.
- Kane, J. M., and J. E. Mertz. (2012). Debunking myths about gender and mathematics performance. *Notices of the AMS*, *59*, 10-21. doi:10.1090/noti790

- Kennedy, J. P., Lyons, T., & Quinn, F. (2014). The continuing decline of science and mathematics enrolments in Australian high schools. *Teaching Science, 60*, 34-46.
- Lareau, A. (2003). *Unequal childhoods: Race, class and family life*. Berkeley: University of California Press.
- Lubienski, S. T., Crane, C. C., & Robinson, J. P. (2011). *A longitudinal study of gender and mathematics using ECLS data*. Final report (grant# R305A080147) submitted to the National Center for Education Research, Institute of Education Sciences, Washington, DC.
- Lubienski, S. T., Robinson, J. P., Crane, C. C., & Ganley, C. M. (2013). Girls' and Boys' Mathematics Achievement, Affect, and Experiences: Findings from ECLS-K. *Journal for Research in Mathematics Education, 44*, 634-645. doi:10.5951/jresmetheduc.44.4.0634
- Lykes, M. B. (2000). Possible contributions of a psychology of liberation: Whither health and human rights?. *Journal of Health Psychology, 5*, 383-397. doi: 10.1177/135910530000500312
- Mack, J. & Wilson, R. (2015). Trends in mathematics and science subject combinations in the NSW HSC 2001-2014 by Gender. Technical paper, University of Sydney. Retrieved from:
<http://www.maths.usyd.edu.au/u/SMS/MMW2015.pdf>
- Mann, A., & DiPrete, T. A. (2016). The consequences of the national math and science performance environment for gender differences in STEM aspiration. *Sociological Science, 3*, 568. doi:10.15195/v3.a25
- Marks, J. L., Lam, C. B., & McHale, S. M. (2009). Family patterns of gender role attitudes. *Sex Roles, 61*, 221-234. doi:10.1007/s11199-009-9619-3
- Martinez Dy, A., Martin, L., & Marlow, S. (2014). Developing a critical realist positional approach to intersectionality. *Journal of Critical Realism, 13*, 447-466. doi: 10.1179/1476743014Z.00000000043
- McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education, 37*, 129-150. Retrieved from
<http://www.jstor.org/stable/30034845>
- Muntaner, C., & Augustinavicius, J. (2019). Intersectionality: A scientific realist critique. *The American Journal of Bioethics, 19*, 39-41. doi: 10.1080/15265161.2018.1557296
- National Innovation and Science Agenda, Australian Government (2017). *National innovation and science agenda*. Retrieved from <http://www.innovation.gov.au/page/agenda>

- National Science Foundation, Division of Science Resources Statistics. (2012). *Women, minorities, and persons with disabilities in science and engineering* (Special Report NSF 11-309). Arlington, VA. Retrieved from <http://www.nsf.gov/statistics/wmpd/sex>
- National Science and Technology Council, Office of Science Technology Policy, US. Government Washington D.C. (2013). https://www.whitehouse.gov/sites/default/files/microsites/ostp/stem_stratplan_2013.pdf
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*, 1067-1101. doi:10.1162/qjec.122.3.1067
- Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *The Journal of Economic Perspectives*, *24*, 129-144. doi:10.1257/jep.24.2.129
- Parker, R., Larkin, T., & Cockburn, J. (2017). A visual analysis of gender bias in contemporary anatomy textbooks. *Social Science & Medicine*, *180*, 106-113. doi: 10.1016/j.socscimed.2017.03.032
- Parker, P. D., Marsh, H. W., Guo, J., Anders, J., Shure, N., & Dicke, T. (2018). An information distortion model of social class differences in math self-concept, intrinsic value, and utility value. *Journal of Educational Psychology*, *110*, 445. doi: 10.1037/edu0000215
- Parker, P. D., Van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction*, *54*, 125-137. doi: 10.1016/j.learninstruc.2017.09.002
- Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology*, *114*, S138-S170. doi:10.1086/589252
- Prilleltensky, I. (2008). The role of power in wellness, oppression, and liberation: The promise of psychopolitical validity. *Journal of Community Psychology*, *36*, 116-136. doi:10.1002/jcop.20225
- Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational and Behavioral Statistics*, *10*, 75-98. doi:10.2307/1164836
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). Newbury Park, CA: Sage.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.

Sáinz, M., & López-Sáez, M. (2010). Gender differences in computer attitudes and the choice of technology-related occupations in a sample of secondary students in Spain. *Computers & Education, 54*, 578-587.

doi:10.1016/j.compedu.2009.09.007

Sinclair, S., Hardin, C. D., & Lowery, B. S. (2006). Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology, 90*, 529. doi:10.1037/0022-3514.90.4.529

Spierings, N. (2012). The inclusion of quantitative techniques and diversity in the mainstream of feminist research. *European Journal of Women's Studies, 19*, 331–347. doi:10.1177/1350506812443621

Sterne, J. A. C., Egger, M., & Moher, D. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1. 0.

Stevenson, H. W., Chen, C., & Booth, J. (1990). Influences of schooling and urban-rural residence on gender differences in cognitive abilities and academic achievement. *Sex Roles, 23*, 535-551.

doi:10.1007/BF00289767

Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PloS one, 11*, e0153857. doi:10.1371/journal.pone.0153857

Stotsky, J., Shibuya, S., Kolovich, L., & Kebhaj, S. (2016). *Trends in women's advancement and gender equality*. International Monetary Fund Working Paper, Washington DC, IMF).

Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement, 63*(5), 765-790.

doi:10.1177/0013164402251027

Van Zanden (2018). Understanding the psychological and social origins of gender disparities in self-beliefs, motivation, and educational attainment. *Unpublished Dissertation*.

von Hippel, P. (2015). The heterogeneity statistic I^2 can be biased in small meta-analyses. *BMC Medical Research Methodology, 15*, 35. doi:10.1186/s12874-015-0024-z

Wenner, G. (2003). Comparing poor, minority elementary students' interest and background in science with that of their white, affluent peers. *Urban Education, 38*, 153-172. doi:10.1177/0042085902250483

Zarrett, N., Malanchuk, O., Davis-Kean, P. E., & Eccles, J. (2006). Examining the gender gap in IT by race: Young adults' decisions to pursue an IT career. In *Women and information technology: Research on underrepresentation*. Eds. J. McGrath and W. Asprey. MIT Press: Cambridge, Massachusetts, London

Zurbriggen, E. L., & Capdevila, R. (2010). The personal and the political are feminist: Exploring the relationships among feminism, psychology, and political life. *Psychology of Women Quarterly*, 34, 458-459.

doi:10.1111/j.1471-6402.2010.01595.x

Table 1

Results of Gender and Expectancy Value Meta-Analyses Across Math, Science, and Verbal Domains

<i>Variable</i>	<i>ESs</i>	<i>D</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t_2</i>	<i>t_3</i>	<i>I²_2</i>	<i>I²_3</i>	<i>Q Statistic</i>
Math expectancy for success	150	0.27	0.23	0.31	-	0.04	-	0.92	2046.31
Math task value	44	0.14	0.06	0.21	-	0.05	-	0.93	567.35
Math intrinsic value	79	0.17	0.12	0.22	-	0.04	-	0.91	1020.50
Math utility value	60	0.08	0.02	0.13	-	0.03	-	0.89	563.94
Math attainment value	8	0.02	-0.10	0.15	-	0.02	-	0.84	23.92
Math cost*	4	0.08	0.21	0.05	-	0.00	-	0.00	3.22
Science expectancy for success	58	0.18	0.10	0.26	0.02	0.04	0.33	0.60	794.65
Science task value	28	0.01	-0.06	0.08	-	0.03	-	0.88	424.73
Science intrinsic value	39	0.21	0.11	0.32	0.02	0.04	0.31	0.62	575.22
Science utility value	16	0.05	-0.02	0.12	-	0.01	-	0.80	141.87
Science attainment value	10	0.05	-0.02	0.13	-	0.01	-	0.64	30.61
Computing expectancy for success	22	0.44	0.28	0.60	-	0.13	-	0.97	198.76
Computing task value	9	0.22	0.07	0.38	-	0.04	-	0.91	30.44
Computing intrinsic value	14	0.48	0.26	0.69	-	0.13	-	0.97	81.37
Computing utility value	10	0.21	0.07	0.35	-	0.02	-	0.86	21.02

(continued)

<i>Variable</i>	<i>ESs</i>	<i>D</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t_2</i>	<i>t_3</i>	<i>I²_2</i>	<i>I²_3</i>	<i>Q Statistic</i>
Engineering expectancy for success*	7	0.24	0.08	0.40	-	0.02	-	0.81	10.54
Engineering intrinsic value*	7	0.22	0.11	0.32	-	0.00	-	0.00	2.03
Engineering utility value*	7	0.04	-0.14	0.22	-	0.01	-	0.78	18.77
Physical sciences expectancy for success	16	0.43	0.29	0.56	-	0.05	-	0.93	60.00
Physical sciences task value*	3	-0.14	-0.65	0.38	-	0.17	-	0.98	13.63
Physical sciences intrinsic value*	8	0.27	0.19	0.36	-	0.00	-	0.25	8.05
Physical sciences utility value*	7	0.05	-0.11	0.21	-	0.01	-	0.68	10.69
Physical sciences cost*	3	-0.32	-0.43	-0.21	-	0.00	-	0.00	5.04
Biological sciences expectancy for success*	5	-0.03	-0.19	0.14	-	0.01	-	0.60	5.92
Biological sciences intrinsic value*	7	-0.23	-0.40	-0.06	-	0.03	-	0.90	26.30
Biological sciences utility value*	4	-0.09	-0.30	0.12	-	0.01	-	0.78	5.64
Verbal expectancy for success	65	-0.17	-0.23	-0.11	-	0.04	-	0.92	353.17
Verbal task value	22	-0.48	-0.62	-0.34	-	0.10	-	0.96	315.92
Verbal intrinsic value	35	-0.32	-0.40	-0.24	-	0.04	-	0.90	196.98
Verbal utility value*	7	-0.27	-0.31	-0.23	-	0.00	-	0.00	7.92
Verbal attainment value	9	-0.28	-0.35	-0.22	-	0.00	-	0.42	13.58

Notes. *d* = Cohen's *d*; *ESs* = number of effect sizes; *CI* = confidence intervals; see text for descriptions of heterogeneity and homogeneity measures. * Note I^2 should be interpreted with caution in small meta-analyses where the number of effects is considered 'small' (e.g., ~ 7 or under). See von Hippel (2015) for a discussion of I^2 biases in small meta-analyses.

Table 2a
Meta-Analyses and Moderation Analyses for Math

<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
Math expectancy for success		149	0.28	0.24	0.32	0.04		0.92	2050.30
Social class	<i>p</i> = 0.00						0.32		
Low SES		25	0.19	0.10	0.28	0.02		0.65	56.87
Middle SES		19	0.25	0.17	0.33	0.00		0.15	23.97
High SES		6	0.77	0.31	1.24	0.31		0.94	85.09
Reliability	<i>p</i> = 0.01						0.08		
<.70 or not reported		55	0.19	0.14	0.24	0.02		0.71	215.39
≥.70		93	0.32	0.26	0.37	0.05		0.94	1598.16
Math task value		44	0.14	0.06	0.21	0.05		0.93	567.35
Social class	<i>p</i> = 0.00						0.65		
Low SES		8	0.03	-0.06	0.12	0.00		0.00	7.29
High SES		5	0.67	0.34	1.00	0.11		0.85	36.13
Math intrinsic value		79	0.17	0.12	0.22	0.04		0.91	1020.50
Social class	<i>p</i> = 0.02						0.38		
Low SES		6	0.05	-0.01	0.11	0.00		0.00	13.17
Middle SES		17	0.08	-0.03	0.19	0.02		0.54	26.75
High SES		5	0.47	0.19	0.75	0.08		0.80	23.22
National Gender Equality	<i>p</i> = 0.03						0.20		
Highest level of gender equality		13	0.32	0.18	0.45	0.05		0.94	181.78
<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
High level of gender equality		19	0.17	0.11	0.23	0.01		0.64	51.93
Medium level of gender equality		39	0.13	0.05	0.21	0.05		0.91	522.59
Low level of gender equality		5	0.04	-0.02	0.11	0.00		0.46	14.23
Math utility value		60	0.08	0.02	0.13	0.03		0.89	567.23
Social class	<i>p</i> = 0.03						1.00		
Low SES		14	-0.05	-0.14	0.04	0.00		0.00	13.77
Middle SES		4	0.10	0.01	0.20	0.00		0.00	2.28
Era	<i>p</i> = 0.02						0.29		
< 1980s		14	0.18	0.06	0.30	0.04		0.72	49.09

1980s	8	0.20	0.03	0.37	0.04		0.87	47.84
1990s	17	0.09	-0.02	0.19	0.02		0.72	95.84
2000s	14	-0.03	-0.10	0.05	0.01		0.86	174.28
Current	7	0.02	-0.10	0.15	0.02		0.77	23.43
% Ethnic minority		$p = 0.03$					0.33	
Lowest	21	0.13	0.03	0.23	0.04		0.85	67.24
Low	4	-0.09	-0.25	0.07	0.02		0.91	128.71
Highest	9	-0.16	-0.24	-0.07	0.00		0.00	12.74

Note. Non-significant results are available in Supplementary Materials. Math utility confidence interval for social class was calculated using a fixed level meta-analysis.

Table 2b
Moderation Analyses for Science

<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
Science expectancy for success		58	0.18	0.10	0.26	0.04		0.60	794.65
Social class	<i>p</i> = 0.01						0.64		
Low SES		7	0.14	0.01	0.27	0.00		0.13	6.41
Middle SES		4	0.56	0.27	0.86	0.07		0.83	26.44
Age	<i>p</i> = 0.03						0.17		
Elementary age		4	-0.05	-0.14	0.03	0.00		0.00	2.51
Middle school age		22	0.09	-0.00	0.18	0.04		0.92	455.49
High school age		25	0.24	0.14	0.33	0.05		0.95	184.42
Young adult		5	0.31	0.11	0.51	0.04		0.81	47.01
National Gender Equality	<i>p</i> = 0.00						0.27		
High level of gender equality		17	0.24	0.12	0.36	0.06		0.94	190.19
Medium level of gender equality		32	0.15	0.08	0.23	0.03		0.86	190.22
Low level of gender equality		6	-0.11	-0.22	-0.00	0.01		0.82	44.33
Gender Segregation in Graduates	<i>p</i> = 0.01						0.33		
Low levels of gender segregation		4	-0.06	-0.27	0.15	0.04		0.91	73.45
Medium levels of gender segregation		28	0.11	0.02	0.20	0.04		0.90	186.55
High levels of gender segregation		6	0.30	0.23	0.37	0.01		0.72	66.48

(continued)

<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
Science task value		28	0.01	-0.06	0.08	0.03		0.88	424.73
National Gender Equality	<i>p</i> = 0.00						0.72		
High level of gender equality		7	0.14	0.12	0.16	0.00		0.00	13.51
Medium level of gender equality		14	-0.02	-0.12	0.07	0.02		0.78	104.72
Low level of gender equality		6	-0.16	-0.19	-0.13	0.00		0.00	9.84
Gender Segregation in Graduates	<i>p</i> = 0.01						0.52		
Medium levels of gender segregation		15	-0.05	-0.14	0.04	0.02		0.81	100.20
High levels of gender segregation		4	0.15	0.12	0.18	0.00		0.10	4.96

Note. Non-significant results are available in Supplementary Materials. Science value confidence interval for high gender equality was calculated using a fixed level meta-analysis.

Table 2c
Moderation Analyses for Computing Sciences

<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
Computing expectancy for success		22	0.44	0.28	0.60	0.13		0.97	198.76
Sample Type	<i>p</i> = 0.02						0.27		
Special sample (e.g., elective STEM)		13	0.61	0.39	0.83	0.12		0.92	82.58
Standard sample		9	0.23	0.05	0.41	0.06		0.95	84.51

Note. Non-significant results are available in Supplementary Materials.

Table 2d
Moderation Analyses for Verbal

<i>Variable</i>	<i>ANOVA p-value</i>	<i>ESs</i>	<i>d</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	<i>t²</i>	<i>R²</i>	<i>I²</i>	<i>Q Statistic</i>
Verbal intrinsic value		35	-0.32	-0.40	-0.24	0.04		0.90	196.98
Era	<i>p</i> = 0.02						0.33		
1990s		18	-0.40	-0.51	-0.30	0.01		0.65	34.99
2000s		7	-0.43	-0.51	-0.36	0.00		0.00	6.66
Current		10	-0.21	-0.34	-0.08	0.04		0.88	76.50
Gender Segregation in Graduates	<i>p</i> = 0.04						0.65		
Medium levels of gender segregation		7	-0.25	-0.40	-0.10	0.01		0.78	14.68
High levels of gender segregation		6	-0.52	-0.67	-0.37	0.01		0.80	9.42
Highest levels of gender segregation		9	-0.36	-0.46	-0.25	0.00		0.00	8.80
Reliability	<i>p</i> = 0.01						0.37		
<.70 or not reported		25	-0.39	-0.47	-0.31	0.02		0.74	65.84
≥.70		10	-0.19	-0.31	-0.07	0.03		0.88	47.67

Note. Non-significant results are available in Supplementary Materials. Verbal intrinsic value confidence interval for 2000s was calculated using a fixed level meta-analysis.

Table 3a
Composite Gender Equality Indexes Moderators of Gender Differences

Outcome	Moderator	ESs	Beta	<i>Lower</i> 95% CI	<i>Upper</i> 95% CI	z	p-value	t ²	R ²	I ²	<i>Q</i> Statistic
Math Expectancy	GGI	146	0.05	0.01	0.09	2.6	0.01	0.04	0.11	96.29	1939.80
Math Expectancy	GII	139	0.04	0.00	0.08	2.27	0.02	0.04	0.07	97.46	1927.34
Math Value	GGI	44	0.06	0.01	0.12	2.38	0.02	0.04	0.21	88.19	567.35
Math Value	GII	42	0.06	0.00	0.12	2.05	0.04	0.04	0.19	91.35	566.27
Math Intrinsic Value	GGI	75	0.04	0.00	0.08	2.1	0.04	0.04	0.1	94.32	985.72
Math Intrinsic Value	GII	70	0.05	0.01	0.09	2.49	0.01	0.04	0.15	97.54	974.35
Science Expectancy	GGI	51	0.08	0.04	0.12	3.83	>.001	0.02	0.33	91.21	602.70
Science Expectancy	GII	50	0.09	0.05	0.13	4.42	>.001	0.02	0.4	96.01	602.00
Science Value	GGI	27	0.06	0.03	0.1	3.39	>.001	0.01	0.43	88.43	414.81
Science Value	GII	27	0.08	0.05	0.11	5.34	>.001	0.01	0.67	94.21	414.81
Science Intrinsic Value	GII	31	0.05	0.01	0.09	2.66	0.01	0.01	0.35	93.67	379.28
Computing Expectancy Physical Science Expectancy	GII	20	-0.14	-0.25	-0.03	-2.55	0.01	0.07	0.41	80.34	142.45
Expectancy	GGI	16	0.11	0.02	0.21	2.32	0.02	0.03	0.4	21.66	60.00

Notes. All Moderators are z-standardized. Beta represents the change in gender differences (in Cohen's D) for a 1 standard deviation change in the moderator. GGI = Gender Gap Index; GII = Gender Inequality Index (reverse coded so higher scores equal more. Equality).

Table 3b

Domain Specific Gender Equality Indexes Moderators of Gender Differences

Outcome	Moderator	ESs	Beta	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>	z	p-value	t ²	R ²	I ²	<i>Q Statistic</i>
Math Expectancy	LR	149	0.04	0.01	0.08	2.44	0.02	0.04	0.09	92.99	2040.51
Math Value	LR	44	0.04	0.00	0.08	2.05	0.04	0.04	0.15	76.03	567.35
Math Utility Value	Prim	59	-0.08	-0.12	-0.03	-3.39	>.001	0.02	0.27	97.87	563.69
Math Utility Value	Parl	56	0.07	0.01	0.14	2.22	0.03	0.03	0.15	93.02	415.76
Science Expectancy	Tert	52	0.05	0.00	0.1	2.17	0.03	0.03	0.12	98.13	695.56
Science Expectancy	LR	52	0.05	0.03	0.08	3.88	>.001	0.03	0.32	79.01	695.56
Science Value	Parl	27	0.12	0.07	0.17	4.62	>.001	0.01	0.6	92.77	414.81
Science Value	LR	27	0.04	0.01	0.06	3.07	>.001	0.01	0.39	90.12	414.81
Science Intrinsic Value	Prim	32	-0.06	-0.10	-0.01	-2.21	0.03	0.01	0.27	95.56	383.31
Science Intrinsic Value	LR	33	0.03	0.01	0.06	2.48	0.01	0.02	0.24	91.37	451.69
Science Utility Value	Prim	15	-0.09	-0.16	-0.02	-2.69	0.01	0	0.6	90.23	133.02
Science Utility Value	Tert	15	-0.11	-0.21	-0.02	-2.27	0.02	0.01	0.46	90.23	133.02
Physical Science Expectancy	Parl	16	0.11	0.01	0.21	2.15	0.03	0.03	0.35	80.00	60.00
Physical Science Expectancy	LR	16	0.34	0.09	0.59	2.64	0.01	0.02	0.52	31.67	60.00

Notes. All Moderators are z-standardized. Beta represents the change in gender differences (in Cohen's D) for a 1 standard deviation change in the moderator. Prim. = male/female primary enrollment ratio; Sec. = male/female secondary enrollment ratio; Tert. = male/female tertiary enrollment ratio; LR = male/female labor force participation ratio; HLMP = women's share of higher labor market positions; WR = women's share of research positions; Parl. = women's share of parliamentary seats.