

Journal of Educational Psychology

Gender Differences in Collaborative Problem-Solving Skills in a Cross-Country Perspective

Francesca Borgonovi, Seong Won Han, and Samuel Greiff

Online First Publication, April 27, 2023. <https://dx.doi.org/10.1037/edu0000788>

CITATION

Borgonovi, F., Han, S. W., & Greiff, S. (2023, April 27). Gender Differences in Collaborative Problem-Solving Skills in a Cross-Country Perspective. *Journal of Educational Psychology*. Advance online publication. <https://dx.doi.org/10.1037/edu0000788>

Gender Differences in Collaborative Problem-Solving Skills in a Cross-Country Perspective

Francesca Borgonovi^{1, 2}, Seong Won Han³, and Samuel Greiff⁴

¹Institute of Education, University College London

²OECD Centre for Skills, OECD, Paris, France

³Department of Educational Leadership and Policy, University at Buffalo

⁴Department of Behavioral and Cognitive Sciences, University of Luxembourg

Effective collaborative problem solving comprises cognitive dimensions, in which men tend to outperform women, and social dimensions in which women tend to outperform men. We extend research on between-country differences in gender gaps by considering collaborative problem solving and its association with two indicators of societal-level gender inequality. The first indicator reflects women's underrepresentation in the labor market and politics. The second reflects women's underrepresentation in stereotypically masculine fields and men's underrepresentation in stereotypically feminine fields among university students. We use cross-country evidence on collaborative problem-solving skills among 15-year-old students from 44 countries ($N = 343,326$) who participated in the 2015 Programme for International Student Assessment (PISA). Girls outperform boys in collaborative problem solving in all countries. Gender gaps in collaborative problem solving in favor of girls are less pronounced in countries where women are especially underrepresented in the labor market and politics but more pronounced in countries where men and women are more likely to conform to gender stereotypes in selecting a field of study at university. Societal-level gender equality plays a bigger role in explaining between-country differences in achievement in domains with a gender gap in favor of girls—such as collaborative problem solving and, to a lesser extent, reading—and a smaller role in explaining between-country differences in achievement in domains with a gender gap in favor of boys—such as mathematics.

Educational Impact and Implications Statement

This work examines gender gaps in collaborative problem solving in the teenage years in the 2015 PISA. Girls outperform boys in collaborative problem solving in all countries in the study. At the country level, the size of the gender gap depends on societal-level gender equality, although associations depend on how societal-level gender equality is operationalized. The gender gap in favor of girls is wider in countries in which women have more equal opportunities compared with men in the labor market and politics. In contrast, the gender gap in favor of girls is smaller in countries in which men are more likely to enroll in fields of study at university where men are underrepresented and women are more likely to enroll in fields in which they are underrepresented—that is, where, for example, women are more likely to enroll in engineering and men are more likely to enroll in teaching programs. Societal-level gender equality measures explain gender gaps in domains in which girls outperform boys—such as problem solving and reading—more than in domains in which boys outperform girls—such as mathematics.

Keywords: collaborative problem solving, gender gaps, gender inequality, Programme for International Student Assessment, cross-country

Supplemental materials: <https://doi.org/10.1037/edu0000788.supp>

Francesca Borgonovi  <https://orcid.org/0000-0002-6759-4515>

Seong Won Han  <https://orcid.org/0000-0002-3676-2483>

Samuel Greiff  <https://orcid.org/0000-0003-2900-3734>

Francesca Borgonovi is employed by the Organisation for Economic Cooperation and Development (OECD), which is responsible for the development of the PISA study used in this work. The opinions expressed do not necessarily reflect the opinions of the OECD or its member countries. Francesca Borgonovi acknowledges support from the British Academy through its Global Professorship program.

Francesca Borgonovi served as lead for conceptualization, project administration, and writing—original draft and served in a supporting role for formal analysis. Seong Won Han served as lead for data curation, formal analysis,

and visualization and served in a supporting role for project administration and writing—original draft. Samuel Greiff served in a supporting role for conceptualization, methodology, and writing—original draft. Francesca Borgonovi and Seong Won Han contributed to methodology and writing—review and editing equally.

Open Access funding was provided by University College London: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; <http://creativecommons.org/licenses/by/4.0>). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Correspondence concerning this article should be addressed to Francesca Borgonovi, UCL Social Research Institute, Institute of Education University College London, 55-59 Gordon Square, London WC1H 0NU, United Kingdom. Email: F.borgonovi@ucl.ac.uk

Effective collaboration depends on the capacity and willingness to solve problems and achieve a set of goals, and to do so by working with others (Graesser et al., 2018). A growing body of literature in psychology and economics assesses gender differences in preferences for collaboration and competition in work and everyday life (Bertrand, 2011; Croson & Gneezy, 2009; Falk & Hermle, 2018; Van Vugt et al., 2007). Such literature generally relies on task experiments and identifies women as more likely to engage in positive social reciprocity, less likely to engage in negative social reciprocity (e.g., Eagly & Wood, 1991; Falk & Hermle, 2018), responding less positively to competitive environments (Niederle & Vesterlund, 2007), being less aggressive (Eagly & Steffen, 1986), more risk adverse (Eckel & Füllbrunn, 2015; Fisk & Ridgeway, 2018), and more empathetic (Christov-Moore et al., 2014). However, the literature has so far focused almost exclusively on adult populations and willingness to collaborate. Descriptive evidence on gender differences in collaborative problem-solving skills among adolescents exists (OECD, 2017a) and previous work examined the role of contextual factors in shaping differences across genders in domain-general problem solving (Borgonovi & Greiff, 2020). However, to date, no study has examined the role of contextual factors in shaping gender differences in collaborative problem solving in adolescence. This is the gap we fill with our contribution.

We provide new empirical evidence on gender differences among 15-year-old students in the ability to engage effectively in goal-directed collaboration and consider whether the between-country variation in the gender gap in collaborative problem solving is related to measures of societal-level gender equality. Furthermore, we consider if the role of societal-level gender equality in shaping gender differences in collaborative problem solving is more or less pronounced than its role in shaping gender differences in academic subjects such as reading and mathematics. We thus extend previous work examining gender differences among teenage populations in domain-general problem solving (Borgonovi & Greiff, 2020) to a setting in which problems are solved via social interactions and in which achieving one's goal requires effective collaboration with others. We also build upon the large literature on gender differences in skills and preferences in curricular subjects among children as well as emerging evidence on non-curricular domains, including financial literacy (Hasler & Lusardi, 2017) and digital literacy (Siddiq & Scherer, 2019).

A crucial contribution of our study is that, in order to identify the role of societal-level gender equality in shaping the gender gap in collaborative problem solving, we use both an indicator of whether women are equally represented in the labor market and politics as well as an indicator of whether both men and women are equally represented in different fields of study at university. We argue that the former more strongly reflects women's empowerment while the latter more strongly reflects the social context experienced by both men and women. Finally, we compare the association between societal-level gender equality and the gender gap in collaborative problem solving with associations in reading—a domain in which, at age 15, the gender gap in favor of girls is very large in virtually all countries—and in mathematics—a domain in which the gender gap in favor of boys is quantitatively smaller than the gap observed for reading and can be observed only in a subset of countries (32 out of the 79 education systems that participated in PISA in 2018; OECD, 2019a).

Past Research and Hypotheses

Collaborative Problem Solving

Collaborative problem solving is broadly defined as “the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution, and pooling their knowledge, skills and efforts to reach that solution” (OECD, 2017a, p. 47). It describes the set of skills individuals need to possess to integrate social and cognitive components when a group of people jointly work on solving a problem (Fiore et al., 2017, 2018).

Theoretical work on collaborative problem solving relies on related fields, such as collaborative learning and collaborative decision making (Graesser et al., 2018), but also differs from them in the sense that collaborative problem solving particularly focuses on situations in which groups encounter new problems that require new solutions and cannot be solved through standard operations. Hence, collaborative problem solving describes the process in which a group and its individual members engage when confronted with a problem situation that they, for a variety of potential reasons, need to solve together (Greiff et al., 2013).

There are different conceptualizations of collaborative problem solving (for an overview, see Graesser et al., 2018). However, virtually all existing frameworks distinguish between social (i.e., collaborative) and cognitive (i.e., problem solving) components. This distinction is especially relevant for our work, since boys and girls may not perform equally well on average on the two dimensions. The conceptualization of the social components mainly draws on research findings on collaboration and group interaction that are partly rooted in social psychology. For instance, this research examines the role of group members' personality (e.g., Barrick et al., 1998) or social cohesion processes (e.g., Beal et al., 2003) that take place during collaborative problem solving. The cognitive components of collaborative problem solving mainly draw on cognitive and action theories that examine the cognitive processes taking place during problem solving (Funke, 2010; Greiff et al., 2013; Mayer & Wittrock, 2006). These theories consider processes such as defining a problem and applying appropriate operators to reduce the difference between the current and goal state (OECD, 2010).

Beyond these similarities, contemporary frameworks for collaborative problem solving differ in two respects: firstly, with regard to which specific social and cognitive processes they consider relevant, and secondly, how these components interrelate with each other. In this article, we rely on the PISA framework, which defines collaborative problem solving along four problem solving and three collaboration dimensions (OECD, 2017a, 2017b). The four problem-solving processes are: (a) *exploring and understanding*; (b) *representing and formulating*; (c) *planning and executing*; and (d) *monitoring and reflecting*. The three collaboration processes are: (a) *establishing and maintaining shared understanding*; (b) *taking appropriate actions to solve the problem*; and (c) *establishing and maintaining group organization*. The PISA 2015 framework was designed to address all combinations of processes (four problem solving and three collaboration processes; 12 cells in total) equally, thus ensuring the framework broadly covers the underlying theoretical concept.

Gender Differences in Collaborative Problem Solving

Collaborative problem solving extends the cognitive requirements necessary for successful problem solving in any domain—such as attention, the mental capacity to store and manipulate different types of information simultaneously, and the capacity to represent and manipulate knowledge structures (Wiley & Jarosz, 2012)—by including additional social and emotional factors that allow individuals to work effectively with others on specified tasks.

Gender differences in collaborative problem solving could arise because of differences across genders in skills, interests, and affects that are relevant for collaborative problem-solving performance and that could, in turn, be shaped by societal-level gender equality. The literature has identified significant gender disparities in interest and affect (Wang & Degol, 2017; Xie et al., 2015). Of particular relevance for our work, previous analyses of PISA data reveal marked gender differences in attitudes toward collaboration among 15-year-old students: on average girls reported valuing relationships more than boys did, whereas boys reported valuing teamwork more than girls (OECD, 2017a). Both students who valued relationships with others and those who valued teamwork performed better in collaborative problem solving than those who valued them less. Indeed, the literature suggests that girls are more likely to perceive themselves as empathic, communal, and cooperative, while boys are more likely to see themselves as agentic (Wood & Eagly, 2012). Furthermore, in social surveys, women are often described as communal and caring, and there is evidence that, at least in the United States, the attribution of communal traits to women rather than men has increased over time (Eagly et al., 2020).

The pursuit of communal rather than agentic goals has been studied alongside differences in math skills and aptitude as factors shaping gender differences in education and employment in science, technology, engineering, and mathematics (STEM) fields—fields in which men are overrepresented (Diekmann et al., 2010, 2011; Evans & Diekmann, 2009; Sczesny et al., 2019; Wood & Eagly, 2012)—but could also, in turn, shape gender differences in domains such as collaborative problem solving. Previous studies suggest that women outperform men in terms of willingness to collaborate (Croson & Gneezy, 2009; Van Vugt et al., 2007). Moreover, a number of studies have hypothesized and empirically found that men tend to develop conceptions of the self based on characteristics such as independence and autonomy, while women are more likely to define the self in terms of relatedness and interdependence, although gender differences in self-construals vary across countries (Cross & Madson, 1997; Watkins et al., 2003; Yang & Girgus, 2019). To the extent that women conceptualize themselves in terms of relationships, they should be particularly motivated to engage in, maintain, and rely upon social connections (Amanatullah et al., 2008; Cross et al., 2000). As a result, gender differences in self-construals could lead girls and women to seek greater engagement with others and, as a result, develop greater skills in the social dimensions of collaborative problem solving.

In contrast, a large international study based on data from the PISA 2012 cycle from over 30 countries worldwide found that boys tend to outperform girls on the cognitive dimensions of problem solving (Borgonovi & Greiff, 2020), another important prerequisite for success in collaborative problem solving. Gender gaps in the cognitive components of problem solving have been ascribed to gender differences in favor of boys in abstract

information processing (Halpern & LaMay, 2000) and spatial and navigation abilities (Baron-Cohen & Wheelwright 2004; Coutrot et al., 2018; Lawton & Hatcher, 2005; Reilly & Neumann, 2013). Differences in the cognitive components of problem solving may have repercussions for gender differences in educational and career choices (Stoet & Geary, 2018).

The size of the gender gap in collaborative problem solving among teenagers is likely to result from boys' and girls' relative advantages in the different factors that lead to proficiency in collaborative problem solving, namely, the cognitive and social dimensions. Given the evidence for contrasting factors shaping overall achievement in collaborative problem solving, it is not possible to hypothesize a priori whether boys will outperform girls or girls will outperform boys in collaborative problem solving.

Sociocultural Factors and Gender Differences in Collaborative Problem Solving

Social science research has proposed different theories that could account for the emergence of gender differences in different skills and attitudes. Existing theories generally consider biological, psychological, and environmental factors as well as the interaction between them (Miller & Halpern, 2014; Wood & Eagly, 2012). On average, differences in cognitive ability, personality, social behaviors, and psychological well-being within each gender are considerably larger than differences between them (Hyde, 2014), but even small differences can be consequential for educational choices, labor market, and broader social outcomes. Moreover, between-gender differences vary across countries. This has led to the examination of sociocultural factors, of which societal-level gender equality is a key sociocultural factor in this research strand (see Keller et al., 2021; Parker et al., 2020 for recent work).

Numerous studies have investigated factors determining gender differences in attitudes, educational and occupational expectations, and subject-specific self-beliefs among students (Correll, 2001; Olsson & Martiny, 2018; Wang & Degol, 2017). For example, educational psychologists have investigated the degree to which gender-incongruent role models (e.g., female engineers and scientists or male teachers) can reduce gender stereotypes and promote counter-stereotypical aspirations and behaviors, such as STEM educational and career aspirations among girls (Cheryan et al., 2011; González-Pérez et al., 2020; Sonnert et al., 2007) or teaching career aspirations among boys (Han et al., 2020). One of the ways in which societal-level gender equality could shape gender gaps in collaborative problem solving is by influencing the roles men and women currently play in society.

According to social role theory, the roles men and women currently play could influence the expectation that future generations of men and women will be suited (or not) to occupy similar (or different) roles if existing differences are ascribed to underlying differences in internal predispositions and ability, that is, if existing differences lead to the development of gender stereotypes (Ashmore & Del Boca, 1979; Eagly, 1987; Eagly & Wood, 2012; Ellemers, 2018; Eccles, 1994; Wood & Eagly, 2012). The theory predicts that in countries where men and women have, on average, similar roles and opportunities in the labor market and in society—that is, countries with higher levels of gender equality—boys and girls will be more likely to expect to be able to play a wide range of roles. As a result, they will be more likely to strive to develop the wide set of

skills and dispositions that will enable them to occupy such roles. In other words, and in line with expectancy value theory (Eccles et al., 1983; Eccles & Wigfield, 2020; Wigfield & Eccles, 2020), greater gender equality at the societal level could promote similar levels of motivation and effort by both boys and girls to acquire a range of skills, rather than develop a narrow set of skills that match those currently used by men (for boys) or by women (for girls). Expectancy value theory in fact stipulates that motivation depends on performance expectancies and task value and in countries with greater gender equality the performance expectancies and task value of boys and girls could be more similar than those of boys and girls living in countries with less gender equality.

Previous empirical evidence indicates that teenage boys' tendency to outperform girls in the cognitive component of problem solving is, on average, larger in countries with lower gender equality, when this is expressed in terms of the degree to which women have similar employment and political opportunities as men (Borgonovi & Greiff, 2020). This evidence matches findings from the broader literature identifying wider gender gaps in skills in which boys tend to outperform girls—such as mathematics—in the presence of lower societal-level gender equality (Breda et al., 2018; Else-Quest et al., 2010; Guiso et al., 2008).

However, gender differences in reading and text comprehension in favor of girls tend to be wider in countries that score high on gender equality indicators that operationalize equality in terms of women's representation in the labor market and politics (OECD, 2015). These findings align with research suggesting that contexts that promote girls' performance in domains in which they are typically disadvantaged also tend to promote their performance in domains in which they are typically advantaged (Reardon et al., 2019).

Gender equality indicators typically employed in the literature, such as the Gender Inequality Index (GII), reflect the degree to which women have reached parity with men in labor force participation, earnings and in the political life but they contain little information about the opportunities men have and the opportunities and barriers they face. Although such indicators can be effective measures of whether a society is able to reduce girls' disadvantage with respect to gender gaps in favor of boys, they do not necessarily reflect the degree to which a society has reduced stereotypes regarding what boys and men are able to achieve and are suited to engage in (Charles & Bradley, 2009). This is arguably more consequential in domains in which gender gaps are in favor of girls/women. There is evidence that although women have made gains in entering fields in which they were traditionally underrepresented, men's progress in entering fields in which they were traditionally underrepresented has been slower (Friedman, 2015), and less attention has been devoted to studying this (Han et al., 2020). As such, it might be important to complement the widely used gender equality indicators in research on gender gaps in education that reflect women's empowerment with indicators that reflect the context experienced by both men and women.

The adoption of such indicators is especially relevant for our work since successful collaborative problem solving requires proficiency in dimensions in which, according to the previous literature, boys tend to outperform girls (cognitive dimensions) but also dimensions in which girls tend to outperform boys (social dimensions). To the extent that societal-level gender equality is conceptualized in terms of increased parity in the expectations and ambitions of both boys and girls, in countries with high levels of gender equality boys' advantage in the

cognitive dimensions of collaborative problem solving should be smaller, as should girls' advantage in the social dimensions, leading to a small overall combined gender gap in collaborative problem solving, the direction of which depends on the relative importance of the different dimensions. In countries with low levels of gender equality, boys' advantage in the cognitive dimensions should be large, but so should girls' advantage in the social dimensions, leading to an overall combined gender gap in collaborative problem solving that will also depend on the relative impact of the social context on the different dimensions. In contrast, wider gender gaps in favor of girls are expected in countries scoring highly on indicators of gender equality that reflect solely women's empowerment, because in these countries, girls' disadvantage in the cognitive dimensions can be expected to be small and their advantage in the social dimensions can be expected to be large. We formulate two alternative hypotheses regarding the role of societal-level gender equality and the size of the gender gap in collaborative problem solving depending on how gender equality is conceptualized and measured.

In the past few years, extensive research and policy efforts have been undertaken to promote women's participation in fields typically considered masculine and to build girls' confidence in their ability in math (see OECD, 2015, 2017d for reviews). In contrast, despite increasing awareness of boys' underachievement in other areas (Borgonovi & Han, 2021; Legewie & DiPrete, 2012, 2014; van Hek et al., 2019), less attention has been given to reducing gender gaps that favor girls. As a result, progress in reducing women's underrepresentation in traditionally masculine fields such as STEM has not been matched by similar progress in reducing men's underrepresentation in traditionally feminine fields such as teaching and nursing (Friedman, 2015). Therefore, we expect that societal-level gender equality will play a stronger role in shaping the achievement of girls than of boys and that the association between societal-level gender equality and the gender gap in achievement will be stronger in domains in which girls outperform boys than in domains in which boys outperform girls.

The Present Study

First, we replicate previous descriptive estimates of the gender gap in collaborative problem solving developed by the OECD (OECD, 2017a) by estimating country-specific gender gaps in collaborative problem solving before but also after controlling for possible confounders. Second, we consider differences in the association between the size of the gender gap in collaborative problem solving and two indicators of societal-level gender (in)equality context. The first indicator is the widely used GII, which measures quantitative differences in women's participation in the economic and political life of a country. We use the GII to test hypotheses on the size of the gender gap in collaborative problem solving when societies empower women since the indicator expresses the extent to which women are held back in society. The second is the Sex Segregation Index (SSI), which measures women's representation in male-dominated fields of study and men's representation in female-dominated fields of study at the tertiary level (Charles & Bradley, 2009; see the "Methods" section for more details). Social scientists have labelled academic fields and occupations dominated by one gender as *masculine* or *feminine* depending on the prevalence of men and women in these fields and the extent to which they require the performance of tasks or use of skills that are stereotypically characterized as corresponding to either males or females' aptitude and preferences

(Correll, 2001; Nosek et al., 2002). Mathematical tasks are often stereotyped as masculine, and as a result, math-intensive fields such as STEM are often viewed as masculine. In contrast, fields that require caring for others, especially the young and sick, such as teaching and nursing, are often viewed as feminine. We use the SSI to test hypotheses on the size of the gender gap in collaborative problem solving when societies empower both men and women, since the SSI indicator considers the extent to which both women and men are present in fields which are counter-stereotypical.

Third, we compare the strength of the associations between the two indicators of societal-level gender equality and the gender gap in collaborative problem solving with the strength of associations between societal-level gender equality indicators and gender gaps in mathematics and reading.

We formulate three sets of hypotheses:

Hypothesis 1: Adolescent boys and girls do not perform equally well in collaborative problem solving.

Hypothesis 2a: The gender gap in favor of girls in collaborative problem solving will be larger in societies that empower women in the political and economic life.

Hypothesis 2b: Contrasting effects lead to uncertain predictions of the relation between the size of the gender gap in collaborative problem solving in societies that empower both men and women to enter counter-stereotypical fields.

Hypothesis 3: Societal-level gender equality is more strongly associated with the achievement of girls than of boys and the association between societal-level gender equality and the gender gap in achievement is stronger in academic domains in which girls outperform boys than in domains in which boys outperform girls.

Data and Method

Participants

All cases used in our analyses were extracted from the public-use files for PISA 2015, available at <https://www.oecd.org/pisa/data/2015database/>. PISA participants were selected from the population of 15-year-old students in each participating country according to a two-stage random sampling procedure, so that the weighted samples were representative of students enrolled in grade 7 or above and between 15 years and 3 months and 16 years and 2 months of age at the time of the assessment administration (generally referred to as 15-year-olds in this work). In the first stage, a stratified sample of schools was drawn. In the second stage, students were selected at random within each sampled school. While 70 educational systems participated in PISA 2015, our study is based on the subset of countries that administered the computer-based assessment for collaborative problem solving. Furthermore, since our analyses aim to identify the association between country-level characteristics and collaborative problem solving, our sample is restricted to the subset of countries for which we were able to identify country-level information. Out of the 70 countries that participated in PISA in 2015, 50 countries administered the collaborative problem-solving assessment alongside the assessments in key academic domains. Out of these 50 countries, six were eliminated from the analyses due to missing data on country-level information. Our analytic sample includes 343,326 students in 44 countries.

Procedures

On the day of the test, students who were selected to take part in the PISA study sat in a dedicated room equipped with computers under the supervision of a test administrator. Participants were first administered a timed 2-hr test and then a questionnaire designed to take around 30 min to complete. Participants were typically selected from different classes and grades. Students first familiarized themselves with the PISA computer platform. They were told that the test would last for 2 hr, with a break after the first hour of testing, and that the test would be followed by a questionnaire. They were also given an opportunity to practice all response formats and to explore the (simple) navigation tools embedded in the test platform before starting the test. After the 2-hr test, students were asked to complete a questionnaire (whose total duration never exceeded 1 hr).

Measures

Descriptive statistics for all variables used in the analyses are presented in Table 1. Note that we provide descriptive statistics on the original scales in Table 1, but standardized continuous variables for the analyses. Correlations between variables are presented in Table 2. It should be noted that although the PISA CPS assessment measures a unique set of abilities, it is also highly correlated with

Table 1
Descriptive Statistics

Variable	<i>M</i>	<i>SD</i>	Min	Max
Individual level				
Collaborative problem-solving (CPS) assessment scores ^a	478.94	100.96	43.25	941.19
ESCS ^d	-0.20	1.10	-7.05	4.07
Girl ^b	0.51		0.00	1.00
Immigrant ^b	0.11		0.00	1.00
Other language at home ^b	0.10		0.00	1.00
PISA math assessment scores ^a	466.07	100.32	70.80	868.64
PISA reading assessment scores ^a	471.72	102.96	0.00	882.12
School level				
(Pre)vocational school ^b	0.23		0.00	1.00
School level ESCS ^d	-0.24	0.79	-4.05	1.58
Urban ^b	0.33		0.00	1.00
National level				
PISA sample selectivity	85.55	9.61	61.71	97.28
GDP per capita	32,894.16	17,313.7	11,436	98,460
GII ^d	0.17	0.11	0.04	0.41
Age of selection	1.41	1.72	0	6
Standardization ^d	0.45	0.25	0.07	0.97
Sex segregation index ^c	2.33	0.44	1.51	3.6

Note. We provide descriptive statistics in the original metric for achievement domains. However, in the analyses continuous variables were grand-mean centered. ESCS = economic, social, and cultural status; PISA = 2015 Programme for International Student Assessment; GDP = gross domestic product; GII = Gender Inequality Index.

^a Achievement scores were standardized to have an *M* of 0 and an *SD* of 1 across analytic sample countries in our work. ^b Since the gender indicator (girl), immigrant, other language at home, (pre)vocational school, and urban variables are dichotomous the mean of these variables indicates the percentage of the sample with value 1. ^c Sex segregation index variable is available only for 30 countries in our analytic sample. The SSI was standardized to have an *M* of 1 and an *SD* of 1 across analytic sample of countries in our work. ^d Continuous measures—ESCS, school mean ESCS, PISA sample selectivity, log of GDP per capita, GII, standardization, age of selection—were standardized to have an *M* of 0 and an *SD* of 1 across the analytic sample.

Table 2
Correlations Between Variables

Variable	CPS	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Student level															
1. Gender	.144*	—													
2. ESCS	.342*	-.038*	—												
3. Immigrant	-.011*	-.016*	.068*	—											
4. Other language at home	-.073*	-.025*	.038*	.455*	—										
5. Math scores	.749*	-.003*	.408*	-.002	-.041*	—									
6. Reading scores	.787*	.150*	.374*	-.021*	-.088*	.852*	—								
School level															
7. (Pre)vocational	-.022*							—							
8. School-level ESCS	.420*						.029*	-.015*	—						
9. Urban	.100*						-.015*	.187*		—					
National level															
10. PISA sample selectivity	.196*										—				
11. Log of GDP per capita	.281*										.538*	—			
12. Gender Inequality Index (GII)	-.317*										-.639*	-.731*	—		
13. Age of selection	-.022*										.071	.117	-.100	—	
14. Standardization	-.156*										-.304*	-.309*	.278*	.030	—
15. Sex Segregation Index	.220*										.436*	.533*	-.440*	-.003	-.497*

Note. CPS = collaborative problem solving; ESCS = economic, social, and cultural status; PISA = 2015 Programme for International Student Assessment; GDP = gross domestic product. * $p \leq .05$.

other measures of achievement in PISA, such as reading and mathematics ($\rho = 0.787$ and $\rho = 0.749$, respectively).

Dependent Variable: Collaborative Problem Solving

The key dependent variable employed in this study was students' skill level in collaborative problem solving as measured in the 2015 edition of PISA. As a convention, PISA scales all of its assessments to have an M of 500 and an SD of 100 across all OECD countries. However, for the purpose of the present study, we rescaled the PISA collaborative problem-solving scores to have an M of 0 and an SD of 1 across the countries making up the analytic sample. PISA 2015 contained a set of 10 plausible values for collaborative problem solving, which we use in the analysis. Plausible values are random draws from the marginal posterior of the latent distribution of ability for each student, in other words, plausible values are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual based on their observed responses in the PISA test.

The assessment instrument was developed by an international consortium with expertise in both the methodological topics associated with the assessment and the substantive field of collaborative problem solving (for more information on the development process, see OECD, 2017b). In addition, the instruments were developed based on the theoretical framework outlined above. Thus, the final instrument reflected the theoretical framework comprised of the 12 cells that define collaborative problem solving.

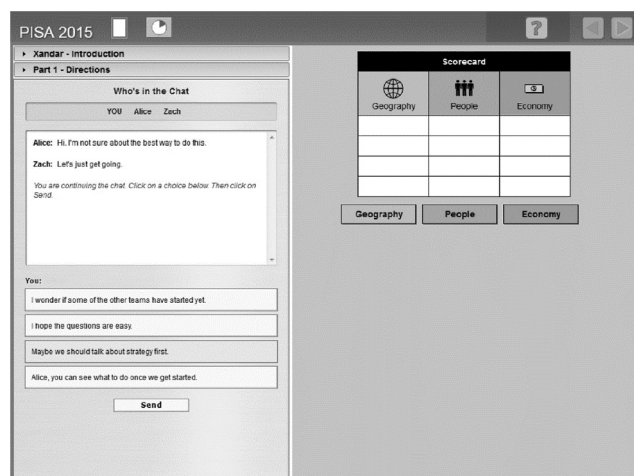
One relevant feature of the PISA 2015 assessment instrument for collaborative problem solving was that, when sitting for the assessment, students did not interact with real peers, but instead with one or more computer-simulated agents. As such, students were confronted with a range of theoretically relevant peer behaviors (e.g., different levels of social cohesion, different personalities among the computer-simulated agents, or different levels of hierarchy in the group; Graesser et al., 2018), which were administered to students in a standardized way. Thus, during the assessment, students interacted with computer-mediated agents via predefined messages in a chat box, and subsequently or simultaneously (depending on the specific stimulus) solved a problem. Students were aware that they were interacting with computer-mediated agents.

A total of six units were developed within the PISA framework. Each unit typically lasted between 5 and 20 min and was comprised of several items. Note that each item was assigned to one of the 12 cells in the theoretical framework, which implied that each item primarily addressed both one collaboration and one problem-solving process. According to the official report on collaborative problem solving (OECD, 2017a), the units were designed in a way that required students to engage in different types of collaboration, including jigsaw or hidden-profile tasks, consensus-building tasks, and negotiation tasks. The rationale behind this was to present students with a variety of different situations typical of real-world scenarios that 15-year-old students would encounter across the globe and that avoid putting any particular subgroup of students at a distinct advantage (e.g., boys or girls).

A sample unit that was published to demonstrate the assessment principle is Xandar (OECD, 2017a). In Xandar, students participate in an in-class contest in which they need to answer quizzes about the fictitious country Xandar. The student is assigned to a group with two computer-mediated agents, Alice and Zach. Throughout the unit, the group of three needs to answer questions on different aspects of Xandar, such as its geography, its people, and its government. To

Figure 1

Screenshot of the Published Sample Unit Xandar (OECD, 2017a) With the Chat Space on the Left Side and the Task Space on the Right Side



find the correct answers, the group together engages in problem solving at different levels of complexity as they work through the unit's items. Figure 1 is a screenshot from Xandar. The figure displays both the chat space (on the left) and the task space (on the right). In the chat space, the student communicates and discusses with Alice and Zach by selecting from a range of predefined messages. In the task space, students can engage in various problem-solving activities, such as tracking their progress or viewing notes. More information, including a detailed description of all the items in Xandar, can be found in OECD (2017a).

Key Independent Variables

Gender is a key individual-level independent variable (IV). This variable was reported by students in the background questionnaire. In all models, we report differences in outcomes associated with being a girl compared to being a boy.

We measure societal-level gender inequality using two indices: the GII and the SSI in different fields of study in higher education. The GII was developed by the United Nations Development Programme (UNDP) and reflects gender inequality in three aspects of human development—reproductive health, social and political empowerment, and economic empowerment. The GII was normalized to have an M of 0 and an SD of 1 across the analytic sample of countries, with higher values indicating greater gender inequality. The SSI measures the degree to which women or men are overrepresented in different fields of study in higher education in each country (Charles & Bradley, 2009). The SSI was normalized so as to have an M of 0 and an SD of 1 across the analytic sample of countries, with higher scores indicating higher gender segregation and therefore greater inequality. As shown in Table 2, the correlation between the GII and the SSI was moderately negative, suggesting that the two indicators capture different aspects of gender inequality ($\gamma = -0.440$).

Controls

We include individual-, school-, and national-level characteristics in our models. The inclusion of these controls is driven by the fact that

boys and girls with different background characteristics may not be equally likely to be part of the PISA target population in different countries and that such differences may be both driven by societal-level gender equality and associated with achievement (Parker, et al., 2020).

At the individual level, we control for: students' economic, social and cultural status (ESCS), whether the student has an immigrant background, whether the language the student speaks at home matches the language of instruction, and the student's performance in reading and mathematics. The ESCS index is an aggregate indicator that reflects students' economic, social, and cultural status and is based on students' answers to items in the PISA background questionnaire asking them to report their parents' educational attainment, occupation, and the availability of a range of resources within their home (OECD, 2017c). The index was normalized ($M = 0$ and $SD = 1$). The literature indicates that socio-economic background is one of the strongest determinants of achievement differences in PISA (Pokropek et al., 2015); moreover, depending on the level of gender equality, socio-economically disadvantaged boys and girls may be differently likely to still be in school at age 15. We introduce a dichotomous indicator that takes a value of 1 if the student reports that the language he or she speaks most frequently at home is different from the language of the PISA test and a value of 0 if it is the same language. Similarly, we introduce a dichotomous indicator that takes a value of 1 if students reported being born in a country other than the country in which they took the PISA test or reported having foreign-born parents, and a value of 0 otherwise. Previous studies show that native- and non-native-speaking immigrant-origin students differ in their participation in education and attainment (Borgonovi & Ferrara, 2020; Buchmann & Parrado, 2006; Dronkers & Kornder, 2014; Suárez-Orozco et al., 2008).

Most of our results controlled for students' achievement in reading and mathematics to show the unique contribution of societal-level gender equality to explain the gender gap in collaborative problem solving, net of gender differences in other academic skills. The only set of models where we omit curricular achievement controls are those in which we compare the strength of the association between societal-level gender equality and the gender gap in collaborative problem solving with the associations between societal-level gender equality and the gender gap in reading and mathematics (Hypothesis 3). As indicators of academic achievement, we use the PISA reading and mathematics achievement scores. We rescale the PISA reading and mathematics achievement scales (which in the original metric have an M of 500 and an SD of 100 across OECD countries), so that each has an M of 0 and an SD of 1 across the countries making up the analytic sample. As shown in Table 2, the correlations between collaborative problem solving, mathematics and reading scores range from 0.749 to 0.832.

At the school level, we control for three factors: the academic orientation of the school or track in which the student is enrolled, whether the school is located in an urban or rural context, and the socio-economic composition of the students attending the school, because there is evidence that gender differences in achievement differ depending on school factors (Legewie & DiPrete, 2012, 2014; OECD, 2015). We control for academic orientation because education systems differ greatly in the prevalence of vocational or pre-vocational programs in upper secondary school (OECD, 2021). In general, participation in such programs is associated with lower achievement, and boys are more likely than girls to participate in vocational and pre-vocational programs (OECD, 2015). Information on academic orientation was obtained through the student tracking form, which indicates whether

the curricular content of the program in which the student was enrolled was general, pre-vocational, or vocational. We introduce a dichotomous variable taking a value of 1 if the student was enrolled in a program with a prevocational or vocational orientation and 0 if the program was general. Degree of urbanicity was reported by school principals. School urbanicity was coded as 1 if principals reported that their school was located in a community with more than 100,000 inhabitants and 0 otherwise. We also take into account the socio-economic composition of the school using an indicator of mean ESCS. Although the school-level ESCS indicator is correlated with individual-level ESCS, the two can be estimated accurately and are routinely used in empirical research of PISA data and in the OECD's own reports (see, e.g., Agasisti et al., 2021; OECD, 2019b). Urbanicity and school-level socio-economic conditions reflect the conditions students experience and the communities in which they live. As such, they represent potentially important confounders of gender gap estimates if boys and girls are not equally likely to attend advantaged or disadvantaged schools and if the context shapes gender gaps in achievement. Note that we do not include other school-level control variables, such as the percentage of students who are immigrants and the percentage of students who speak a language other than the language of the PISA test at home, since these are highly associated with a school's socio-economic conditions and our objective is simply to control for confounders.¹

Four country-level variables were used as control variables: (a) gross domestic product (GDP) per capita based on purchasing power parity (PPP; in current U.S. dollars); (b) PISA sample selectivity; (c) educational standardization; and (d) early selection of education systems. We used the GDP per capita indicator available through the World Bank Open Data portal (<https://data.worldbank.org/>). Since gender equality tends to be higher on average in more prosperous countries, it is important to control for GDP to net out any potential effects of economic prosperity on overall levels of achievement and gender gaps in achievement.

Controlling for PISA sample selectivity allows us to make meaningful comparisons across countries where the PISA test covers different shares of the 15-year-old population and where out-of-school populations might differ by gender (Han et al., 2018). PISA contains representative samples of 15-year-old students enrolled in educational institutions at the lower secondary school level or above. Results may reflect sample selectivity in the PISA survey to the extent that different numbers of young people in this age group in each country had dropped out of school or were still in primary education—groups that may be particularly low-achievers. While a majority of OECD member countries have achieved near-universal access to schooling, in some countries in PISA 2015, <80% of 15-year-olds were enrolled in school and were thus eligible to participate in PISA (OECD, 2016). This implies that PISA results for some countries are not fully representative of their 15-year-old populations, and such differential representation may differ by gender, with a possible bearing on estimates of gender differences in collaborative problem solving. To take into account these between-country differences in PISA sample eligibility and reduce the possibility that PISA sample eligibility influences the interpretations of PISA results, we calculated the PISA sample selectivity using the share of the weighted number of PISA participating students in the total population of 15-year-olds. Results are similar when countries with the largest levels of sample selectivity are excluded from the analysis. We control for features of education systems because prior research has shown that features of education systems are related to gender inequality in educational outcomes as well as gender

differences in attitudes (Van de Werfhorst & Mijs, 2010; van Hek et al., 2019). At the country level, the level of standardization of the educational system is indicated by the degree to which school curricula are nationally or regionally standardized. We constructed this using the PISA school questionnaire. PISA asked school principals: "Regarding your school, who has considerable responsibility for the following tasks: (a) choosing which textbooks are used, (b) determining which courses are offered, and (c) determining course content. We took the percentage of school principals who reported that local/regional or national educational authorities were responsible. The average score of these three measures indicates the level of standardization in a country; a higher score refers to a higher level of standardization. The level of early selection of each education system was measured by the age of first selection into different school types or tracks. The country with the highest level of early selection is Austria, which divides students into different educational tracks starting at age 10. Countries with no early selection are assigned the age at which students leave secondary education, typically age 16. We subtract the age of first selection in each country from 16, so that it ranges between zero and six, with higher values indicating earlier selection.

Analytical Method

All analyses in the study were conducted using the statistical software Stata 15 and taken into account missing values through multiple imputations by chained equation (MICE; Royston & White, 2011). We generated 20 imputed datasets. The imputation model includes all the variables used in the analyses, as well as socio-demographic variables and student performance in reading, science, mathematics, and collaborative problem-solving skills. Because multiple imputation by chained equations does not automatically accommodate incomplete level-2 variables, imputations were performed for all student-level and school-level characteristics, respectively. Fixed effects at the country level were included in the imputation models to account for potential specificities of individual countries (c.f., Lüdtke et al., 2017).

PISA test scores are based on item response theory and are comparable across students taking different test forms. Because in PISA 2015 a set of 10 plausible values were reported, we combined these, as per OECD recommendations (OECD, 2017c). That is, all analyses were undertaken 10 times, once with each relevant plausible value variable. The results were then averaged, and then significance tests adjusting for variation between the 10 sets of results were computed (details available in PISA 2015 technical report; OECD, 2017c).

In a first step of the analysis, we report descriptive statistics on gender differences in collaborative problem-solving proficiency in each of the 44 countries in our sample. We report results from two sets of models: the observed gender gap and the gender gap while controlling for individual and school-level factors to test Hypothesis 1 on the direction and pervasiveness of the gender gap in collaborative problem solving. In Table S2 in the online supplemental materials, we further report the variance among boys and among girls as well as the variance ratio to determine whether the variability in collaborative solving is higher among males or females.

¹ We ran preliminary models including the percentage of students who are immigrants and percentage of students who speak a language other than the language of the PISA test at home. The coefficients of both variables were close to zero and adding these two models did not improve AIC.

Results in this first step were obtained by estimating linear regression models for each country, before and after controlling for individual- and school-level variables. In this set of models, we use balanced repeated replication weights (BRR) to take into account the clustered nature of the PISA data (students nested in schools) and obtain unbiased estimates for *SEs*. We preferred to use BRR weights rather than running two-level hierarchical models in this first step because BRR weights are preferable when estimating country-specific models in PISA since they account for the specificities of each country's sample (two-stage sampling stratified by public/private school type, for example) and not just school-level clustering like in the multilevel setting. Estimates were obtained by combining 10 sets of results because PISA reading, math, and collaborative problem-solving scores were included (Little & Rubin, 1987; OECD, 2017c).

In a second step, we develop three-level hierarchical linear models where students (level 1) are nested within schools (level 2) and countries (level 3). Recent evidence suggests that using only level two weights, in our case final school weights, is preferable in two-level multilevel modeling (students nested within school models; Mang et al., 2021). However, in this study, since we have a three-level hierarchical model, we use normalized student final weights such that the sum of weights was equal to the number of students in the dataset, and each country contributed equally to the analysis (OECD, 2017c; Rutkowski et al., 2010). We tested if the choice of sampling weights affects main findings and Table S4 in the online supplemental materials indicates that findings are consistent when we used normalized final student weights and school weights as recommended by Mang et al. (2021).

In the baseline null model, we estimate the intraclass correlations (ICCs) in collaborative problem solving to identify how much of the variance lies at the level of individuals, schools, and countries by estimating the empty unconditional model as follows:

Level 1 (student)

$$Y_{ijk} = \pi_{0jk} + e_{ijk},$$

where Y_{ijk} is an outcome index for a student i in school j in country k , and e_{ijk} is a random error associated with each student.

Level 2 (school)

$$\pi_{0jk} = \beta_{00k} + r_{0jk}.$$

Level 3 (country)

$$\beta_{00k} = \gamma_{000} + u_{00k}.$$

In order to investigate the association between gender gaps in collaborative problem solving and societal-level gender inequality, we estimated the following conditional model with random intercepts and random slope for gender:

$$\begin{aligned} Y_{ijk} = & \gamma_{000} + \gamma_{001}(\text{PISA sample selectivity}) \\ & + \gamma_{002}(\text{GDP per capita}) + \gamma_{003}(\text{GIISII}) \\ & + \gamma_{004}(\text{Age of selection}) + \gamma_{005}(\text{Standardization}) \\ & + \gamma_{010}(\text{School mean ESCS}) + \gamma_{020}(\text{Urbanicity}) \\ & + \gamma_{030}(\text{[Pre] vocational}) + \gamma_{100}(\text{Girl}) + \gamma_{200}(\text{ESCS}) \\ & + \gamma_{300}(\text{Immigrant}) + \gamma_{400}(\text{Language at home}) \\ & + \gamma_{500}(\text{Math score}) + \gamma_{600}(\text{Reading score}) \\ & + \gamma_{101}(\text{Girl} * \text{GII/SII}) + \gamma_{102}(\text{Girl} \times \text{Age of selection}) \\ & + \gamma_{103}(\text{Girl} \times \text{Standardization}) + e_{ijk} + r_{0jk} + u_{00k} + u_{10k}. \end{aligned}$$

All continuous covariates at the student, school, and country levels were standardized to have an M of 0 and an SD of 1 across the analytic sample and grand mean centered. We estimated a random intercept and random slope in the gender models; all other slopes were fixed. We estimated more complex sets of models in which we tested whether the main parameters of interest varied when we relaxed the constraint of fixed slopes for the control variables at the individual and school level. We report a subset of these models in Table S4 in the online supplemental materials while the rest can be requested from the authors. We report results from four sets of models, which we run twice, once for each of the two alternative indicators of societal-level gender inequality. In Model 1, we include all individual-level, school-level, and country-level controls as well as societal-level gender inequality indices. In Model 2, we examine the association between either GII or SSI and gender gaps in problem solving by introducing the cross-level interaction between being a girl (level 1) and each of the two country-level measures of societal-level gender inequality separately (level 3). In Model 3, we include additional country-level measures related to national education systems as well as the country-level measures of societal-level gender inequality. Models 2 and 3 allow us to test Hypothesis 2a (when we fit the model using the GII) and Hypothesis 2b (when we fit the model using the SSI). Hypothesis 2a was tested using data from 44 countries because the GII was available for 44 countries, while Hypothesis 2b was tested using data from 30 countries because the SSI was available only for 30 countries. For robustness, we run the GII models on the restricted set of countries for which both indicators are available; the results are aligned with those presented for 44 countries, thus confirming that any difference between the two models is not due to the set of countries selected.

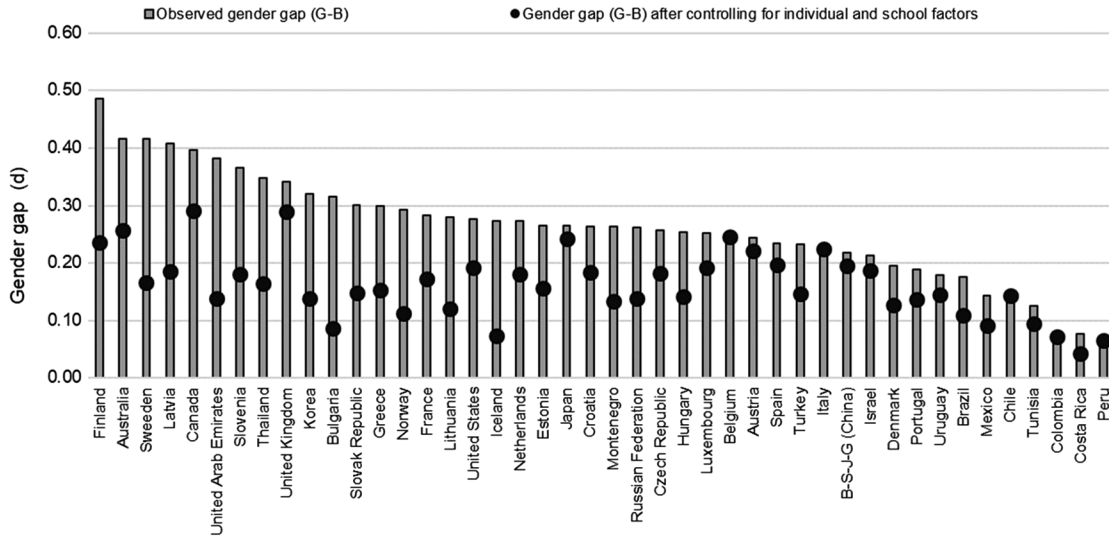
Finally, we test Hypothesis 3 by comparing levels of collaborative problem-solving achievement of girls/boys in countries with high/low level of societal-level equality and by comparing the extent to which the two indicators of societal-level gender inequality are associated with the levels of achievement of boys and girls in collaborative problem solving as well as mathematics and reading. We do this by fitting the Model 3 specification using three outcome indicators—collaborative problem solving, mathematics, and reading—and two societal-level gender inequality indicators—the GII and SSI. Because in this case, we are interested in directly comparing collaborative problem solving, reading, and mathematics, in this specification we do not control for reading and mathematics in the model estimating collaborative problem solving.

Results

Gender Differences in Collaborative Problem Solving

In Figure 2, we report estimates of the gender gap (girls–boys) in collaborative problem solving obtained before but also after controlling for background characteristics (country-specific estimates and associated *SEs* are available in Table S1 in the online supplemental materials). The observed gender gap bars replicate findings reported in the original OECD descriptive reports for collaborative problem solving (OECD, 2017a). The results support the hypothesis that boys and girls do not perform at similar levels in collaborative problem solving (Hypothesis 1), and in fact, indicate that girls tend to

Figure 2
Gender Differences in Collaborative Problem Solving, by Country



Note. Countries are ranked in descending order of the observed gender gap in collaborative problem solving (gray bars). The observed gender gap corresponds to the difference in mean collaborative problem-solving between girls and boys. The observed gender gap is represented using gray bars. All estimates of the observed gender gap are statistically significant at least at the 1% level except for Costa Rica (the only country in which the gender gap is not statistically significant at least at the 5% level after controlling for these individual and school level factors) and Iceland and Colombia (the only countries in which the gender gap is statistically significant at the 5% but not the 1% level). The black dots represent the size of the gender gap after controlling for socio-economic status, immigrant status, language spoken at home, PISA reading and mathematics scores, school program type, school mean socio-economic status, and the location of the school. All estimates account for the nested structure of PISA data. Full estimates are available in Table S1 in the online supplemental materials. PISA = 2015 Programme for International Student Assessment.

outperform boys. In all countries in our sample, girls have higher collaborative problem-solving skills than boys, but the size of the gender gap differs across countries. Comparing the observed gender gap with the gender gap after controlling for gender differences in background characteristics, including reading and math proficiency, suggests that background characteristics explain a large share of the observed gender differences in collaborative problem-solving skills in some countries, but little in others. Overall, gender differences in general cognitive skills and other background characteristics explain around 38% of the gender gap in collaborative problem-solving skills. However, the results also show that a large portion of the gender gap remains unexplained, and crucially, that the role of background characteristics also differs across countries. Table S2 in the online supplemental materials further suggests that although the variability in boys' scores is generally higher than the variability in girls' scores, the differences are minor: the variance ratio is 1.20 or higher only in Finland and South Korea.

The Role of Gender Inequality in Explaining Between-Country Differences in the Gender Gap in Collaborative Problem-Solving Skills

Baseline model estimates reported in Tables 3 and 4 illustrate how much of the variance in collaborative problem solving lies at the student level and how much lies at the school and country levels. The results suggest that around 60% of the overall variance in collaborative problem solving lies within individuals, around 23% can be

attributed to schools, and 16–17% to countries. The $ICC_{country}$ for the baseline model in Table 3 is $\rho = 0.168$ and the $ICC_{country}$ for the baseline model in Table 4 is $\rho = 0.157$. The ICC_{school} for the baseline model in Table 3 is $\rho = 0.227$ and the ICC_{school} for the baseline model in Table 4 is $\rho = 0.225$ (Hedges et al., 2012). The fact that quantitatively meaningful differences exist between countries leads us to explore the role of societal-level factors. Because 44 countries have information on GII (Table 3), whereas only 30 countries have information on SSI (Table 4), the ICCs for the baseline model in Tables 3 and 4 are different because of the different sets of countries included in the analyses.

Tables 3 and 4 present results on the association between key country-level IVs and the gender gap in collaborative problem solving. Table 3 shows the association between GII and collaborative problem solving as well as how the size of the gender gap in collaborative problem solving depends on levels of GII, while Table 4 shows the association between SSI and collaborative problem solving as well as how the size of the gender gap in collaborative problem solving depends on levels of SSI. The results confirm the findings shown in Figure 2 identifying a gender gap in favor of girls in collaborative problem solving, which corresponds to around 16–18% of an *SD* depending on whether we use GII in Table 3 or SSI in Table 4 as the main country-level IV. Furthermore, the models reveal a statistically significant random slope coefficient for the gender covariate at the country level, in line with results presented in Figure 2 that the gender gap varies across countries even when compositional differences are taken into account. Tables 3 and 4 reveal

Table 3*Association Between the Gender Gap in Collaborative Problem Solving and the Gender Inequality Index*

Outcome: Collaborative problem solving	Model 0 (null model)		Model 1		Model 2		Model 3A		Model 3B	
	β	SE	β	SE	β	SE	β	SE	β	SE
Individual level										
Intercept	-0.053	0.065	-0.084***	0.022	-0.084***	0.019	-0.080***	0.017	-0.080***	0.017
Girls			0.157***	0.010	0.155***	0.008	0.155***	0.008	0.156***	0.008
ESCS			0.014*	0.005	0.014*	0.005	0.014*	0.005	0.014*	0.005
Immigrant			-0.024	0.015	-0.024	0.015	-0.024	0.015	-0.024	0.015
Other language at home			-0.052***	0.013	-0.052***	0.013	-0.052***	0.013	-0.052***	0.013
PISA math scores			0.311***	0.013	0.311***	0.013	0.311***	0.013	0.311***	0.013
PISA reading scores			0.413***	0.016	0.413***	0.016	0.413***	0.016	0.413***	0.016
School level										
(Pre)vocational			-0.005	0.014	-0.005	0.014	-0.005	0.014	-0.005	0.014
ESCS mean			0.061***	0.013	0.061***	0.013	0.061	0.013	0.061***	0.013
Urban			0.020*	0.009	0.020*	0.009	0.020*	0.009	0.020*	0.009
National level										
PISA sample selectivity			-0.053	0.034	-0.053	0.030	-0.070*	0.026	-0.070*	0.026
GDP per capita			0.056	0.037	0.056	0.036	0.045	0.028	0.045	0.028
GII			-0.019	0.049	-0.024	0.044	-0.024	0.043	-0.024	0.043
Age of selection							-0.025	0.018	-0.025	0.013
Standardization							-0.058**	0.017	-0.057**	0.018
Cross-level interaction										
Girls \times GII					-0.033**	0.011	-0.033**	0.011	-0.029**	0.010
Girls \times Age of Selection									0.000	0.005
Girls \times Standardization									-0.011	0.007
Random intercept (level 1)	0.618		0.355		0.355		0.352		0.355	
Random intercept (level 2)	0.232***		0.033***		0.033***		0.031***		0.032***	
Random intercept (level 3)	0.172***		0.015***		0.015***		0.011***		0.012***	
Random slope gender (level 3)	—		0.003***		0.003***		0.003***		0.003***	
Proportion of variance explained: Country intercept	—		90.05%		90.08%		92.75%		92.28%	
Proportion of variance explained: Country gender slope	—		62.18%		70.67%		68.58%		70.12%	
AIC	841,777		639,662		639,655		632,413		639,648	

Note. Data are from the PISA 2015 database (<https://www.oecd.org/pisa/data/2015database/>). The dependent variable is the PISA collaborative problem-solving score. Sample size in each specification: 343,326 students, 12,353 schools, and 44 countries. ESCS = economic, social, and cultural status; PISA = 2015 Programme for International Student Assessment; GDP = gross domestic product; GII = Gender Inequality Index; AIC = Akaike information criterion.

*** $p \leq .001$. ** $p \leq .01$. * $p \leq .05$.

that the inclusion of cross-level interactions between the individual level indicator of whether the respondent is a girl and the indicator of societal-level gender inequality yields a lower Akaike information criterion (AIC) but also that the AIC improvement is small.

The results reveal that in countries where women are held back compared with men in economic and political life, the gender gap in collaborative problem solving in favor of girls is smaller. As shown in Model 2, for example, whereas the gender gap in favor of girls corresponds to 0.155 in a country with average levels of GII, it corresponds to 0.122 in countries with greater inequality ($GII = 1$) and 0.188 in countries with lower inequality ($GII = -1$). These results support Hypothesis 2a, in which we hypothesized that the gender gap in favor of girls in collaborative problem solving will be larger when women have greater empowerment. The results are robust to the inclusion of cross-level interaction effects between being a girl and other features of education systems in Model 3B of Table 3.

Results reported in Table 4 indicate that the choice of indicator used to characterize societal-level gender inequality has an important bearing on estimates. In Hypothesis 2b, we formulated an uncertain prediction regarding the size of the gender gap in collaborative

problem solving as a function of the SSI measures, since this measure should affect the gender gap in both the cognitive and social dimensions, but in opposite directions. Whereas the estimates obtained when considering the GII (a measure of women's empowerment) suggest that the gender gap in collaborative problem solving is larger when there is greater equality between men and women in the labor market and politics, the estimates obtained using the SSI (a measure that reflects the segregation of men and women in different fields) suggest that in countries where men and women are less segregated by gender across different fields, the gender gap in favor of girls in collaborative problem solving is smaller. In other words, in societies where gender roles play a more prominent role in shaping male and female university students' field of study (and eventually occupation), disparities in favor of girls in collaborative problem solving are magnified.

The right panel in Figure 3 indicates that the gender gap corresponds to 0.169 in a country with average levels of SSI, but 0.213 in countries with greater inequality ($SSI = 1$) and 0.125 in countries with lower inequality ($SSI = -1$). Although the results are estimated on a restricted sample of countries for which the SSI is available, the fact that the estimates obtained using the GII for the same set of

Table 4*The Association Between the Gender Gap in Collaborative Problem Solving and the Sex Segregation Index*

Outcome: Collaborative problem solving	Model 0 (null model)		Model 1		Model 2		Model 3A		Model 3B	
	β	SE	β	SE	β	SE	β	SE	β	SE
Individual level										
Intercept	0.072	0.066	0.042	0.029	0.042	0.022	0.047*	0.022	0.046*	0.022
Girls			0.169***	0.012	0.169***	0.010	0.169***	0.010	0.170***	0.009
ESCS			0.010	0.006	0.010	0.006	0.010	0.006	0.010	0.006
Immigrant			-0.034	0.017	-0.034*	0.017	-0.034*	0.017	-0.034*	0.017
Other language at home			-0.047**	0.015	-0.047**	0.015	-0.047**	0.015	-0.047**	0.015
Math scores			0.318***	0.015	0.318***	0.014	0.318***	0.014	0.318***	0.014
Reading scores			0.416***	0.020	0.415***	0.019	0.415***	0.019	0.415***	0.019
School level										
(Pre)vocational			-0.007	0.017	-0.007	0.017	-0.007	0.017	-0.007	0.017
ESCS mean			0.056**	0.017	0.056***	0.016	0.056***	0.016	0.056***	0.016
Urban			0.016	0.012	0.016	0.012	0.016	0.012	0.016	0.012
National level										
PISA sample selectivity			-0.044	0.055	-0.044	0.045	-0.056	0.042	-0.057	0.041
Log of GDP per capita			0.079	0.047	0.077	0.040	0.059	0.044	0.059	0.043
Sex segregation index (SSI)			0.046	0.047	0.035	0.031	0.019	0.025	0.019	0.024
Age of selection							-0.027	0.015	-0.025	0.015
Standardization							-0.048*	0.019	-0.047*	0.020
Cross-level interaction										
Girls \times SSI					0.044***	0.010	0.044***	0.010	0.043***	0.011
Girls \times Age of Selection									-0.008	0.008
Girls \times Standardization									-0.001	0.011
Random intercept (level 1)	0.646		0.324		0.324		0.324		0.324	
Random intercept (level 2)	0.236***		0.034***		0.034***		0.034***		0.034***	
Random intercept (level 3)	0.165***		0.004***		0.012***		0.010***		0.010***	
Random slope gender (level 3)	—		0.004***		0.002***		0.002***		0.002***	
Proportion of variance explained:										
Country intercept	—		97.40%		91.09%		92.56%		92.56%	
Country gender slope	—		62.93%		76.68%		76.68%		77.60%	
AIC	572,527		402,389		402,377		402,375		402,379	

Note. ESCS = economic, social, and cultural status; PISA = 2015 Programme for International Student Assessment; AIC = Akaike information criterion. *** $p \leq .001$. ** $p \leq .01$. * $p \leq .05$.

countries are aligned with those reported in the main specification in Table 3 suggests that country selection is not the reason for the discrepancy in results (see Table S3 in the online supplemental materials). The results are robust to the inclusion of cross-level interaction effects between being a girl and other features of education systems in Model 3B of Table 4.

Results presented in Table 3 and Figure 3 also support Hypothesis 3 since they indicate that girls have worse collaborative problem solving in countries where women lag behind men in the economic and political life of their country than in countries in which women have similar opportunities as men and in countries where men and women are more likely to enter counterstereotypical occupations. In contrast, boys have similar collaborative problem solving irrespective of position of women in the economic and political life of their country. Girls appear to be more sensitive to societal level societal level gender inequality while boys are not, irrespective of the indicator used to characterize societal level gender inequality.

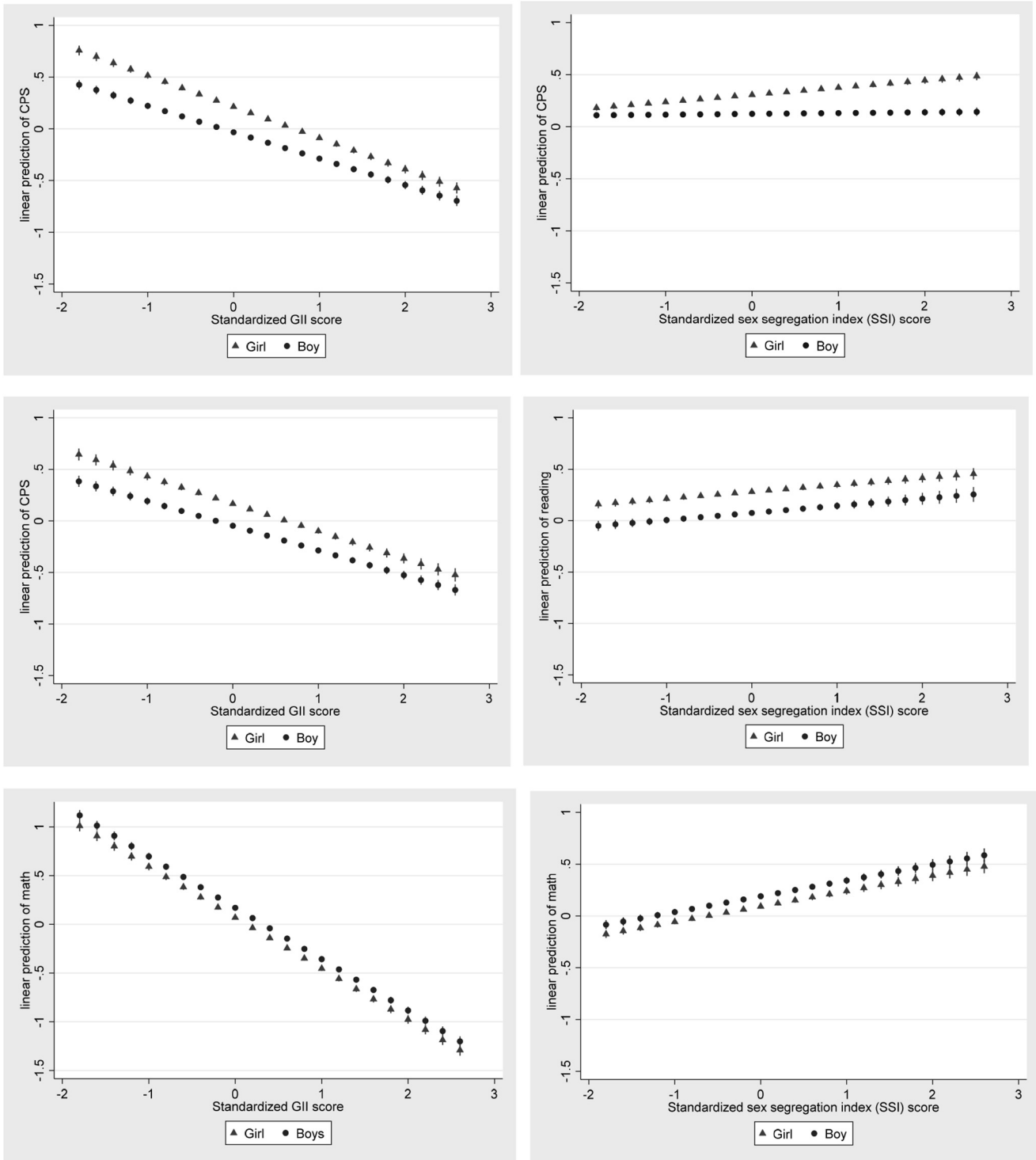
Comparing the Role of Societal-Level Gender Inequality in Explaining Gender Differences in Collaborative Problem Solving, Mathematics, and Reading

The results presented in Table 5 and Figure 3 support Hypothesis 3 in which we hypothesized first that girls would be more responsive

to societal-level gender equality than boys and second that societal-level gender equality is more strongly associated with the gender gap in domains in which girls outperform boys such as collaborative problem solving and reading and lower in domain in which boys outperform girls such as mathematics. In particular, the gender gap in collaborative problem solving is more dependent on the level of gender equality because the achievement of girls in problem solving appears to be strongly associated with societal level gender equality whereas boys' achievement is not. The results illustrate that irrespective of which country-level independent variable is used to characterize gender equality, GII in Panel A at the top and SSI in Panel B at the bottom, the gender gap in collaborative problem solving and reading in favor of girls in countries with average levels of gender equality corresponds to just over one-fifth of an *SD* in collaborative problem solving and reading. The gender gap in mathematics in favor of boys in countries with average levels of gender equality is around half as large, corresponding to around 10% of an *SD*. In countries where women are more likely to be held back in economic and political life, gender gaps in collaborative problem solving and reading are smaller (a 1 *SD* difference in GII corresponds to a weakening of the gender gap in these domains of 0.048 and 0.037 *SD*, respectively). For mathematics, the effects are not statistically significant at conventional levels but are consistent in sign with hypotheses regarding women's empowerment (i.e., wider gaps in favor of

Figure 3

The Gender Gap in Collaborative Problem Solving, Reading, and Math, and How They Vary Depending on the GII and SSI Indices



Note. CPS = collaborative problem solving. All graphs are drawn from Table 5. All continuous covariates are fixed at their mean and results reflect linear predictions of native students who speak at home the same language used in the PISA test. GII = Gender Inequality Index; SSI = Sex Segregation Index; PISA = 2015 Programme for International Student Assessment.

Table 5*The Association Between Societal-Level Gender Inequality and the Gender Gap in Collaborative Problem Solving, Reading and Mathematics*

Independent variables	Collaborative problem solving		Reading		Math	
	β	SE	β	SE	β	SE
Panel A: GII as key country-level IV						
Individual level						
Intercept	-0.069	0.041	-0.067	0.049	0.121*	0.050
Girls	0.212***	0.011	0.214***	0.014	-0.106***	0.010
ESCS	0.120***	0.008	0.145***	0.011	0.150***	0.012
Immigrant	-0.093**	0.031	-0.093*	0.036	-0.100**	0.035
Other language at home	-0.208***	0.018	-0.244***	0.020	-0.179***	0.020
School level						
(Pre)vocational	-0.199***	0.060	-0.263**	0.087	-0.277***	0.083
ESCS mean	0.356***	0.023	0.419***	0.028	0.385***	0.029
Urban	0.042**	0.014	0.046***	0.012	0.013	0.014
National level						
PISA sample selectivity	-0.202***	0.047	-0.162*	0.073	-0.169	0.095
GDP per capita	-0.001	0.077	-0.017	0.066	-0.066	0.082
GII	-0.182*	0.083	-0.136	0.096	-0.257*	0.126
Age of selection	0.028	0.031	0.046	0.043	0.102**	0.036
Standardization	-0.119**	0.037	-0.067	0.040	-0.102**	0.036
Cross-level interaction						
Girls \times GII	-0.048**	0.015	-0.037*	0.017	-0.015	0.013
Girls \times Age of Selection	-0.028**	0.010	-0.038**	0.012	-0.038***	0.008
Girls \times Standardization	-0.022*	0.009	-0.012	0.011	-0.021*	0.009
Random intercept (level 3)	0.058***		0.085***		0.078***	
Random slope gender (level 3)	0.005***		0.008***		0.003***	
Panel B: SSI as key country-level IV						
Individual level						
Intercept	0.056	0.046	0.073	0.052	0.274***	0.056
Girls	0.226***	0.013	0.219***	0.018	-0.109***	0.012
ESCS	0.126***	0.012	0.155***	0.015	0.164***	0.016
Immigrant	-0.113***	0.029	-0.105***	0.030	-0.111***	0.030
Other language at home	-0.204***	0.021	-0.237***	0.023	-0.181***	0.022
School level						
(Pre)vocational	-0.210**	0.076	-0.277*	0.114	-0.283**	0.106
ESCS mean	0.384***	0.030	0.453***	0.035	0.431***	0.033
Urban	0.033*	0.017	0.042**	0.015	0.001	0.016
National level						
PISA sample selectivity	-0.154*	0.067	-0.165	0.100	-0.047	0.114
GDP per capita	0.086	0.071	0.050	0.071	0.039	0.073
SSI	0.097	0.059	0.105	0.063	0.093	0.069
Age of selection	0.005	0.033	0.012	0.041	0.082	0.041
Standardization	-0.092*	0.044	-0.058	0.050	-0.065	0.056
Cross-level interaction						
Girls \times SSI	0.058***	0.016	0.027	0.025	0.013	0.015
Girls \times Age of Selection	-0.034**	0.012	-0.036**	0.013	-0.036***	0.010
Girls \times Standardization	-0.001	0.013	0.009	0.015	-0.016	0.010
Random intercept (level 3)	0.045***		0.060***		0.060***	
Random slope gender (level 3)	0.004***		0.007***		0.003***	

Note. Data are from the PISA 2015 database (<https://www.oecd.org/pisa/data/2015database/>). GII = Gender Inequality Index; SSI = Sex Segregation Index; ESCS = economic, social, and cultural status; PISA = 2015 Programme for International Student Assessment; GDP = gross domestic product.

*** $p \leq .001$. ** $p \leq .01$. * $p \leq .05$.

boys in countries with higher GII). In contrast, in countries where men and women less likely to enroll in fields of study at university in which they are underrepresented, the gender gaps in collaborative problem solving and reading in favor of women are wider (a 1 *SD* difference in SSI corresponds to a weakening of the gender gap in these domains of 0.058 and 0.027 *SD*, respectively). For mathematics, the effects are much smaller in size and indicate that the gender gap in favor of boys is wider in the presence of greater gender equality.

Discussion

Technological advancements, digitalization, and globalization have fundamentally transformed the set of skills needed for workforce readiness and social and personal well-being in the 21st century (Autor et al., 2003; Frank et al., 2019). Mastering a broad set of cognitive skills that are acquired in childhood, such as mathematics, reading, and scientific inquiry, remains crucial, but must now be accompanied by investments in building 21st century skills

throughout the life course (Greiff & Borgonovi, 2022). Skills such as collaborative problem solving are paramount for meeting the demands of a rapidly changing, increasingly dynamic, and unpredictable environment.

Many of the key problems facing societies today are so complex that solving them requires joint effort and cooperation by groups rather than single individuals (Graesser et al., 2022). Teams with diverse expertise and backgrounds are recruited (Hall et al., 2018). An important dimension of diversity is gender. However, efforts to unlock the potential of diverse teams can work only if individuals in such teams possess the capacity to engage in meaningful goal-directed collaboration (Fiore et al., 2018; Graesser et al., 2018; Hesse et al., 2015). Evidence emerging from the analysis of skills requirements in online job vacancies in highly technical fields such as Artificial Intelligence also reveals that the ability to collaborate with others is in high demand and prized alongside the ability to program, develop, and use machine learning algorithms (Samek et al., 2021). By examining the conditions that promote collaborative problem-solving skills among 15-year-old boys and girls in 2015, our work contributes key insights into the readiness of this cohort of individuals to work alongside others as adults. We consider 15-year-olds in 2015 for reasons of data availability; because the teenage years are a period of major neurological and physical changes (Sapolsky, 2017), a process that has the potential to shape individuals' skills, preferences, and attitudes as adults; and because individuals at this age make important educational, training, and labor market decisions. As such, capacity for collaborative problem solving can determine individuals' success in their academic work, in the transition from education to the labor market, and the educational and career choices they make (Diekman & Steinberg, 2013; Evans & Diekman, 2009).

Looking at data from 2015 also allows us to identify the extent to which initial education may fail to equip all young people with the ability to effectively collaborate with others, leading to skills mismatches in the labor market regarding this key competence. Furthermore, our work can guide educational interventions to improve students' learning in the future by providing an in-depth analysis of the role of social context in promoting collaborative problem solving in the recent past. Schools have been increasingly called upon to develop and implement instructional activities that promote learning alongside and working with others (Hmelo-Silver, 2004; Fiore et al., 2018). Our work suggests that considering differences among learners and their sociocultural contexts, rather than applying a one-size-fits-all pedagogical approach, is key to the success of such interventions. In order to be successful problem solvers, individuals need to master a complex array of cognitive skills and must be willing and be able to collaborate with others.

Gender Gaps in Collaborative-Problem Solving and Societal Level Gender Inequality

Proficiency in collaborative problem solving reflects social dimensions as well as cognitive dimensions. Our work highlights that 15-year-old girls tend to outperform boys in collaborative problem-solving skills and that girls are more strongly influenced by the societal-level gender context than boys. On average, the gender gap in collaborative problem solving corresponds to around 16% of an *SD* across the countries in our sample. While such a difference between boys and girls may appear small according to standards first

introduced by Cohen (1988) and Funder and Ozer (2019) suggest that even small differences can have potentially consequential effects. This is especially true for collaborative problem solving, an inherently social activity in which a group's overall achievement may depend on the ability of its weakest link (Kremer, 1993). Therefore, although improvements in model fit were modest when we introduced gender-specific associations between societal-level gender equality and collaborative problem-solving achievement, our results are relevant for education policy and practice. At a general level, our work strengthens the available evidence on the pervasive disparities that exist between genders. It also suggests that the way in which societies are organized today is related to the developmental trajectories of girls and boys, which, in turn, could shape gender disparities in the future. Understanding how classrooms and schools could be organized to facilitate the acquisition of collaborative problem-solving skills among all students could help efforts to redevelop national curricula and learning goals with a stronger focus on 21st century skills such as collaborative problem solving (Graesser et al., 2022). Unfortunately, while there is a considerable body of evidence on the principles underpinning students' development in academic subjects such as text comprehension and mathematics (Hattie & Donoghue, 2016), there is no evidence yet on the effectiveness of such principles and practices in other domains, such as collaborative problem solving (Greiff & Borgonovi, 2022). Our results suggest that it could be important for teachers and educators to be aware of the role of contextual factors in shaping the acquisition of skills among different groups of students, so that they could more easily employ strategies limiting the influence, for example, of stereotypes (Carlana, 2019). Previous work on teacher expectations for female and male students generally focuses on domains such as reading, mathematics, and behavioral aspects (Robinson & Lubienski, 2011; Tiedemann, 2002). The literature indicates that there are differences in teachers' acceptance of different traits and behaviors in boys and girls, but also that the same trait or behavior can be perceived as desirable or undesirable depending on cultural gender role norms (Kerr, 2001). Further research could explore heterogeneity in the sensitivity of different groups of students to the social context and the extent to which such heterogeneity stems from educational interventions in schools, household practices, or individual differences.

How Societal-Level Gender Inequality Measured Can Influence Findings on Gender Gaps in Achievement

Another key finding of our work is that the choice of indicator used to characterize an important feature of the social context in which gender gaps in education arise—societal-level gender equality may have an important bearing on the estimated associations when the focus is on domains in which women/girls outperform men/boys rather than fields in which men/boys outperform women/girls. In recent years, the number of studies empirically estimating the association between societal-level gender equality and the gender gap in academic achievement has grown, and several studies have been published in educational psychology journals (Parker et al., 2020). Most existing empirical work utilizes indicators of women's empowerment—such as the widely used GII, which considers the underrepresentation of women in key life domains—rather than indicators such as the SSI which consider more broadly the opportunity set of both men and women. The SSI, for example,

reflects the underrepresentation of men and women in counter-stereotypical fields of study at university. Whereas predictions from the two sets of indicators align whenever women's disadvantage is the focus of analyses, they diverge whenever male disadvantage is in focus.

Boys' underachievement, especially in domains like literacy for which data are widely available, has started to attract increasing interest among education policymakers and researchers (DiPrete & Buchmann, 2013; Kunnskapsdepartementet, 2019; Legewie & DiPrete, 2012), particularly in light of the emerging (and growing) gender gap in educational attainment in favor of girls. As researchers examine the role of institutional features and societal factors in shaping boys' underachievement (Borgonovi & Han, 2021; van Hek et al., 2019), social context measures that reflect the lived experiences of men as well as women should be developed and employed. Our work represents a first step in this direction.

Previous work has considered the difference between educational conditions that empower girls and educational conditions that enhance the opportunities that both girls and boys have to explain geographical variations in the gender gap in text comprehension and mathematics in the United States (Reardon et al., 2019). We find that when girls grow up in societies that restrict opportunities for women to participate on par with men in economic and political life, the gender gap in favor of girls is lower. In societies where women are more likely to participate on par with men in economic and political life, the gender gap in collaborative problem solving in favor of girls is wider because girls achieve at a higher level compared with when they live in societies where their opportunities are more restricted. In contrast, in societies where men and women on average are more likely to enroll in similar fields of study at university, the gender gap in collaborative problem solving in favor of girls is less pronounced. Our analyses are consistent when we control for achievement in academic subjects like reading and mathematics achievement, suggesting that any association we identify is additional to the association existing between women's representation in the labor market and politics and curricular domains.

Limitations and Future Directions

Although PISA makes it possible to identify relationships cross-nationally, countries decide whether or not to participate (i.e., they self-select), and there are differences in the level of participation in different countries. At the country level, selection into participation is determined by the benefits decision-makers in different countries see as associated with participation in PISA. Such benefits are generally lower in countries in which only a few 15-year-olds are still in school and/or expected achievement levels are very different from the distribution of achievement typical of OECD countries. Country-level participation also depends on countries' ability to comply with the strict technical standards defined by the OECD (2017b) and capacity to bear the high cost of administering the study (OECD, n.d.-a). As a result, at the country level, coverage is skewed toward high-income and upper-middle-income countries, although efforts such as PISA for Development initiative have been implemented to broaden participation in PISA, particularly in light of the use of PISA to measure countries' progress toward meeting the Sustainable Development Goals (OECD, n.d.-b, 2018). In the context of the PISA for Development initiative assessment instruments aligned with the original PISA scales were developed.

These instruments have so far been implemented in Bhutan, Cambodia, Ecuador, Guatemala, Honduras, Panama, Paraguay, Senegal, and Zambia among 15-year-old children enrolled in school as well as out-of-school children. Furthermore, because our study relies on a portion of the test that was administered solely via computer, our sample of countries is even more selected toward high and upper-middle-income countries than the overall PISA country sample. Fifteen countries administered only paper-based instruments because children in these countries could not be expected to be able to work in a computer-based environment or because schools did not have the capacity to set up testing facilities with computers. The 44 countries that had the capacity to administer a complex standardized test on a computer in 2015 participated in the collaborative problem-solving assessment. As a result, our study, like all cross-country analyses based on PISA data, should be interpreted as reflecting associations in the context of mostly high and upper-middle-income countries.

Within each country, the PISA target population is comprised of children between the ages of 15 years 3 months and 16 years 2 months (referred to as 15-year-olds) who are enrolled in lower secondary or upper secondary schools, irrespective of the specific grade they attend. This target population was chosen because in 2000, the year in which PISA was first administered, age 15 was around the age at which compulsory schooling ended in OECD countries, and even accounting grade repetition, the majority of students of this age in OECD countries were enrolled in secondary school. As a result, children who dropped out of school before age 15 or who are still attending primary school are excluded from the PISA target population, making samples not representative of the most disadvantaged 15-year-olds.

Our study focuses on 15-year-old students and, as such, complements work conducted among adults. Further research could examine even younger children in an attempt to identify the developmental trajectory of gender differences in collaborative problem-solving skills and preferences for collaboration.

The PISA 2015 collaborative problem-solving assessment required participants to interact with computer-simulated agents rather than other humans (Graesser et al., 2017), thus posing potential issues with regard to ecological validity. The human to computer agent model was adopted in PISA because it allowed test developers to standardize the instruments and administration conditions and control all assessment parameters. Few empirical studies have examined the extent to which collaborative problem-solving assessments yield different results depending on whether they have participants interact with other human agents or with computer-based agents (Herborn et al., 2020; Stadler et al., 2020; see also Nouri et al., 2017). Studies with samples from Germany and Luxembourg did not identify major differences depending on mode of interaction and indicated that the PISA 2015 assessment was able to tap into relevant aspects of real-world collaborative problem solving.

Finally, our study presents associations rather than causal effects, and as such, should be interpreted as illustrating the extent to which the observed between-country variation in gender disparities in collaborative problem-solving skills varies according to the level of gender inequality present in a society. Moreover, our results are limited to the measure of collaborative problem-solving skills that we used. Future research should ideally validate our findings with other measures, and in particular, could extend our findings by attempting to decompose the overall associations into the

different sets of skills that lead to successful collaborative problem solving.

Conclusion

The results we present in this work provide a basis for a shift in the empirical approach to studying the relationship between societal-level gender equality and gender gaps in education. Future research should consider carefully whether the GII and similar indicators that focus solely on women's empowerment can truly be considered good measures of gender equality, since they fail to consider whether the educational and occupational opportunities of both men and women are influenced by stereotypes. Our work demonstrates the importance of capturing in greater detail the influence of gender roles on skill development and youths' formation of social preferences, while calling into question previous research in the area. Such research has come with the paradoxical result that in societies with greater gender equality (as measured by female participation in the labor market and political life of a country), social preferences are even more polarized between men and women (Falk & Hermle, 2018). This has been interpreted as evidence that the availability of material and social resources creates opportunities for men and women to express gender-specific preferences. In light of our findings, it would be important to re-evaluate such research and examine if the results differ when using alternative indicators that are more reflective of gender social roles and stereotypes.

Finally, we compared the strength of the associations observed for collaborative problem solving with the associations observed for reading (another domain in which girls outperform boys) and mathematics (a domain in which boys outperform girls). This is because much of the literature on societal-level gender equality and academic achievement focuses on mathematics—a precursor of participation in STEM fields—as a way to explain the persistent underrepresentation of women in domains in which men are overrepresented. Our intention was to strengthen the knowledge base regarding the conditions that promote the acquisition of a broader set of competences, going beyond the academic constructs that are the usual focus of educational interventions and curricular decisions. Our findings suggest that the between-country variation in the gender gap in mathematics achievement is smaller than comparable gaps for collaborative problem solving and reading. Furthermore, societal-level gender equality measures explain the between-country variation in the gender gap in collaborative problem solving and reading to a greater degree than mathematics. This could be due to the greater efforts that have been put in place to reduce the gender gap in mathematics over the past few years and the relative paucity of similar policies and interventions targeting boys' underachievement in other domains. Our results could provide the evidence needed to focus political discourse on other dimensions of gender gaps beyond girls' relative underrepresentation among the highest achievers in mathematics.

Findings from our study, if replicated, would have major implications for the interpretation of the role of social context in general and societal-level gender equality in particular for the study of gender gaps in educational achievement. First, they would indicate that analysts need to be mindful of the way in which societal-level gender equality is operationalized. Second, as the labor market and societies increasingly require teamwork and collaboration, our findings call for increased emphasis in schools on equipping youngsters with

collaborative problem-solving skills, with interventions adapted to how pupils' potential is shaped by the sociocultural context in which they live and learn.

References

- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2021). What school factors are associated with the success of socio-economically disadvantaged students? An empirical investigation using PISA data. *Social Indicators Research*, 157(2), 749–781. <https://doi.org/10.1007/s11205-021-02668-w>
- Amanatullah, E. T., Morris, M. W., & Curhan, J. R. (2008). Negotiators who give too much: Unmitigated communion, relational anxieties, and economic costs in distributive and integrative bargaining. *Journal of Personality and Social Psychology*, 95(3), 723–738. <https://doi.org/10.1037/a0012612>
- Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Towards a cognitive-social psychological conceptualization. *Sex Roles*, 5(2), 219–248. <https://doi.org/10.1007/BF00287932>
- Autor, D. H., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83(3), 377–391. <https://doi.org/10.1037/0021-9010.83.3.377>
- Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6), 989–1004. <https://doi.org/10.1037/0021-9010.88.6.989>
- Bertrand, M. (2011). New perspectives on gender. In D. Card, & O. Ashenfelter (Eds.), *Handbook of labor economics* (Vol. 3b, pp. 1543–1590). Elsevier. [https://doi.org/10.1016/S0169-7218\(11\)02415-4](https://doi.org/10.1016/S0169-7218(11)02415-4)
- Borgonovi, F., & Ferrara, A. (2020). The academic and social integration of non-native-speaking immigrant students: The role of linguistic distance. *Learning and Individual Differences*, 81, Article 101911. <https://doi.org/10.1016/j.lindif.2020.101911>
- Borgonovi, F., & Greiff, S. (2020). Societal level gender inequalities amplify gender gaps in problem solving more than in academic disciplines. *Intelligence*, 79, Article 101422. <https://doi.org/10.1016/j.intell.2019.101422>
- Borgonovi, F., & Han, S. W. (2021). Gender disparities in fear of failure among 15-year-old students: The role of gender inequality, the organisation of schooling and economic conditions. *Journal of Adolescence*, 86(1), 28–39. <https://doi.org/10.1016/j.adolescence.2020.11.009>
- Breda, T., Jouini, E., & Napp, C. (2018). Societal inequalities amplify gender gaps in math. *Science*, 359(6381), 1219–1220. <https://doi.org/10.1126/science.aar2307>
- Buchmann, C., & Parrado, E. A. (2006). Educational achievement of immigrant-origin and native students: A comparative analysis informed by institutional theory. In D. P. Baker & A. W. Wiseman (Eds.), *Perspectives on education and society: Vol. 7. The impact of comparative education research on institutional theory* (pp. 335–366). Emerald Group Publishing. [https://doi.org/10.1016/S1479-3679\(06\)07014-9](https://doi.org/10.1016/S1479-3679(06)07014-9)
- Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3), 1163–1224. <https://doi.org/10.1093/qje/qjz008>
- Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, 114(4), 924–976. <https://doi.org/10.1086/595942>

- Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Social Psychological Personality Science*, 2(6), 656–664. <https://doi.org/10.1177/1948550611405218>
- Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience and Biobehavioral Reviews*, 46, 604–627. <https://doi.org/10.1016/j.neubiorev.2014.09.001>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, 106(6), 1691–1730. <https://doi.org/10.1086/321299>
- Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology*, 28(17), 2861–2866. <https://doi.org/10.1016/j.cub.2018.06.009>
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–474. <https://doi.org/10.1257/jel.47.2.448>
- Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of Personality and Social Psychology*, 78(4), 791–808. <https://doi.org/10.1037/0022-3514.78.4.791>
- Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, 122(1), 5–37. <https://doi.org/10.1037/0033-2909.122.1.5>
- Diekmann, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, 21(8), 1051–1057. <https://doi.org/10.1177/0956797610377342>
- Diekmann, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to stem careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology*, 101(5), 902–918. <https://doi.org/10.1037/a0025199>
- Diekmann, A. B., & Steinberg, M. (2013). Navigating social roles in pursuit of important goals: A communal goal congruity account of STEM pursuits. *Social and Personality Psychology Compass*, 7(7), 487–501. <https://doi.org/10.1111/spc3.12042>
- DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation.
- Dronkers, J., & Kornder, N. (2014). Do migrant girls perform better than migrant boys? Deviant gender differences between the reading scores of 15-year-old children of migrants compared to native pupils. *Educational Research and Evaluation*, 20(1), 44–66. <https://doi.org/10.1080/13803611.2013.874298>
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum.
- Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, 75(3), 301–315. <https://doi.org/10.1037/amp0000494>
- Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior. A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100(3), 309–330. <https://doi.org/10.1037/0033-2909.100.3.309>
- Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, 17(3), 306–315. <https://doi.org/10.1177/0146167291173011>
- Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). Sage Publications. <https://doi.org/10.4135/9781446249222.n49>
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly*, 18(4), 585–609. <https://doi.org/10.1111/j.1471-6402.1994.tb01049.x>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.), *Perspective on achievement and achievement motivation* (pp. 75–146). W. H. Freeman.
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, Article 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eckel, C. C., & Füllbrunn, S. C. (2015). Thar she blows? Gender, competition, and bubbles in experimental asset markets. *American Economic Review*, 105(2), 906–920. <https://doi.org/10.1257/aer.20130683>
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, 69, 275–298. <https://doi.org/10.1146/annurev-psych-122216-011719>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127. <https://doi.org/10.1037/a0018053>
- Evans, C. D., & Diekmann, A. B. (2009). On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly*, 33(2), 235–249. <https://doi.org/10.1111/j.1471-6402.2009.01493.x>
- Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412), Article eaas9899. <https://doi.org/10.1126/science.aas9899>
- Fiore, S. M., Graesser, A., & Greiff, S. (2018). Collaborative problem solving education for the 21st century workforce. *Nature: Human Behavior*, 2(6), 367–369. <https://doi.org/10.1038/s41562-018-0363-y>
- Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., Massey, C., O'Neil, H., Pellegrino, J., Rothman, R., Soule, H., & von Davier, A. A. (2017). *Collaborative problem solving. Considerations for the national assessment of educational progress*. National Center for Education Statistics.
- Fisk, S. R., & Ridgeway, C. L. (2018). Framing gender. In B. J. Rismann, C. M. Froyum, & W. J. Scarborough (Eds.), *Handbook of the sociology of gender* (pp. 157–171). Springer.
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14), 6531–6539. <https://doi.org/10.1073/pnas.1900949116>
- Friedman, S. (2015). Still a “stalled revolution”? Work/family experiences, hegemonic masculinity, and moving toward gender equality. *Sociology Compass*, 9(2), 140–155. <https://doi.org/10.1111/soc4.12238>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- González-Pérez, S., Mateos de Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is it a female role-model thing? *Frontiers in Psychology*, 11, Article 2204. <https://doi.org/10.3389/fpsyg.2020.02204>
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. <https://doi.org/10.1177/1529100618808244>
- Graesser, A. C., Kuo, B., & Liao, C. (2017). Complex collaborative solving in assessments of collaborative problem solving. *Journal of Intelligence*, 5(2), Article 10. <https://doi.org/10.3390/jintelligence5020010>
- Graesser, A. C., Sabatini, J. P., & Li, H. (2022). Educational psychology is evolving to accommodate technology, multiple disciplines, and twenty-

- first century skills. *Annual Review of Psychology*, 73(1), 547–574. <https://doi.org/10.1146/annurev-psych-020821-113042>
- Greiff, S., & Borgonovi, F. (2022). Teaching of 21st century skills needs to be informed by psychological research. *Nature Reviews Psychology*, 1(6), 314–315. <https://doi.org/10.1038/s44159-022-00064-w>
- Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving*, 5(2), 71–91. <https://doi.org/10.7771/1932-6246.1153>
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880), 1164–1165. <https://doi.org/10.1126/science.1154094>
- Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P. & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, 73(4), 532–548. <https://doi.org/10.1037/amp0000319>
- Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12(2), 229–246. <https://doi.org/10.1023/A:1009027516424>
- Han, S. W., Borgonovi, F., & Guerriero, S. (2018). What motivates high school students to want to be teachers? The role of salary, working conditions, and societal evaluations about occupations in a comparative perspective. *American Educational Research Journal*, 55(1), 3–39. <https://doi.org/10.3102/0002831217729875>
- Han, S. W., Borgonovi, F., & Guerriero, S. (2020). Why don't more boys want to become teachers? The effect of a gendered profession on students' career expectations. *International Journal of Educational Research*, 103, Article 101645. <https://doi.org/10.1016/j.ijer.2020.101645>
- Hasler, A., & Lusardi, A. (2017). *The gender gap in financial literacy: A global perspective*. Global Financial Literacy Excellence Center. <https://gflc.org/wp-content/uploads/2017/07/The-Gender-Gap-in-Financial-Literacy-A-Global-Perspective-Report.pdf>
- Hattie, J., & Donoghue, G. (2016). Learning strategies: A synthesis and conceptual model. *NPJ Science of Learning*, 1, Article 16013. <https://doi.org/10.1038/npsj.scilearn.2016.13>
- Hedges, L. V., Hedberg, E. C., & Kuyper, A. M. (2012). The variance of intraclass correlations in three- and four-level models. *Educational and Psychological Measurement*, 72(6), 893–909. <https://doi.org/10.1177/0013164412445193>
- Herborn, K., Stadler, M., Mustafic, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015. Can computer agents replace humans? *Computers in Human Behavior*, 104, Article 105624. <https://doi.org/10.1016/j.chb.2018.07.035>
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Springer.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16, 235–266. <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>
- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–398. <https://doi.org/10.1146/annurev-psych-010213-115057>
- Keller, L., Preckel, F., Eccles, J. S., & Brunner, M. (2021). Top-performing math students in 82 countries: An integrative data analysis of gender differences in achievement, achievement profiles, and achievement motivation. *Journal of Educational Psychology*, 114(5), 966–991. <https://doi.org/10.1037/edu0000685>
- Kerr, M. (2001). Culture as a context for temperament: Suggestions from the life courses of shy Swedes and Americans. In T. D. Wachs, & G. A. Kohnstamm (Eds.), *Temperament in context* (pp. 139–152). Lawrence Erlbaum.
- Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3), 551–575. <https://doi.org/10.2307/2118400>
- Kunnskapsdepartementet. (2019). *Nye sjanser—Bedre læring* [New chances—Better learning]. <https://www.regjeringen.no>
- Lawton, C. A., & Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex Roles*, 53, 717–725. <https://doi.org/10.1007/s11199-005-7736-1>
- Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, 77(3), 463–485. <https://doi.org/10.1177/0003122412440802>
- Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education*, 87(4), 259–280. <https://doi.org/10.1177/0038040714547770>
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Lütke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 27(1), 141–165. <https://doi.org/10.1037/met0000096>
- Mang, J., Kuchenhoff, H., Meinck, S., & Prenzel, M. (2021). Sampling weights in multilevel modeling: An investigation using PISA sampling structures. *Large-scale Assessments in Education*, 9(6), Article 6. <https://doi.org/https://doi.org/10.1186/s40536-021-00099-0>
- Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 287–303). Lawrence Erlbaum.
- Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences*, 18(1), 37–45. <https://doi.org/10.1016/j.tics.2013.10.011>
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3), 1067–1101. <https://doi.org/10.1162/qjec.122.3.1067>
- Nosek, B. A., Banaji, M. R., & Greenwald, G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, 83(1), 44–59. <https://doi.org/10.1037/0022-3514.83.1.44>
- Nouri, J., Åkerfeldt, A., Fors, U., & Selander, S. (2017). Assessing collaborative problem solving skills in technology-enhanced learning environments—The PISA framework and modes of communication. *International Journal of Emerging Technologies in Learning*, 12(4), 163–174. <https://doi.org/10.3991/ijet.v12i04.6737>
- OECD. (2010). *PISA 2012 field trial problem solving framework*. <https://www.oecd.org/pisa/pisaproducts/46962005.pdf>
- OECD. (2015). *The ABC of gender equality in education*. <https://doi.org/10.1787/9789264229945-en>
- OECD. (2016). *PISA 2015 Results (volume I): Excellence and equity in education*. <https://doi.org/10.1787/9789264266490-en>
- OECD. (2017a). *PISA 2015 Results (volume V): Collaborative problem solving*. <https://doi.org/10.1787/9789264285521-en>
- OECD. (2017b). *PISA 2015 Assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving*. <https://doi.org/10.1787/9789264281820-en>
- OECD. (2017c). *PISA 2015 technical report*.
- OECD. (2017d). *The pursuit of gender equality: An uphill battle*. <https://doi.org/10.1787/9789264281318-en>
- OECD. (2018). *PISA For development assessment and analytical framework: Reading, mathematics and science*. <https://doi.org/10.1787/9789264305274-en>
- OECD. (2019a). *PISA 2018 Results (volume II): Where all students can succeed*. <https://doi.org/10.1787/b5fd1b8f-en>
- OECD. (2019b). *PISA 2018 Results (volume III): What school life means for students' lives*. <https://doi.org/10.1787/acd78851-en>
- OECD. (2021). *Education at a glance 2021: OECD indicators*. <https://doi.org/10.1787/b35a14e5-en>

- OECD. (n.d.-a). *Requirements for OECD partner countries and economies to participate in PISA*. <https://www.oecd.org/pisa/aboutpisa/howtojoinpisa.htm#:~:text=1%20The%20base%20international%20overhead,and%20reporting%20of%20the%20data>
- OECD. (n.d.-b). *PISA for development*. <https://www.oecd.org/pisa/pisa-for-development/>
- Olsson, M., & Martiny, S. E. (2018). Does exposure to counterstereotypical role models influence girls' and women's gender stereotypes and career choices? A review of social psychological research. *Frontiers in Psychology, 9*, Article 2264. <https://doi.org/10.3389/fpsyg.2018.02264>
- Parker, P. D., Van Zander, B., Marsh, H. W., Owen, K., Duineveld, J. J., & Noetel, M. (2020). The intersection of gender, social class, and cultural context: A meta-analysis. *Education Psychology Review, 32*(1), 197–228. <https://doi.org/10.1007/s10648-019-09493-1>
- Pokropek, A., Borgonovi, F., & Jakubowski, M. (2015). Socio-economic disparities in academic achievement: A comparative analysis of mechanisms and pathways. *Learning and Individual Differences, 42*, 10–18. <https://doi.org/10.1016/j.lindif.2015.07.011>
- Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender achievement gaps in U.S. school districts. *American Educational Research Journal, 56*(6), 2474–2508. <https://doi.org/10.3102/0002831219843824>
- Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles, 68*(9–10), 521–535. <https://doi.org/10.1007/s11199-013-0269-0>
- Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48*(2), 268–302. <https://doi.org/10.3102/0002831210372249>
- Royston, P., & White, I. R. (2011). Multiple imputation of by chained equations (MICE): Implementation in stata. *Journal of Statistical Software, 45*(4), 1–20. <https://doi.org/10.18637/jss.v045.i04>
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary data analysis and reporting. *Educational Researcher, 39*(2), 142–151. <https://doi.org/10.3102/0013189X10363170>
- Samek, L., Squicciarini, M., & Cammeraat, E. (2021). *The human capital behind AI: Jobs and skills demand from online job postings* (OECD Science, Technology and Industry Policy Papers No. 120). OECD Publishing. <https://doi.org/10.1787/2e278150-en>
- Sapolsky, R. M. (2017). *Behave: The biology of humans at our best and worst*. Penguin Press.
- Sczesny, S., Nater, C., & Eagly, A. H. (2019). Agency and communion: Their implications for gender stereotypes and gender identities. In A. E. Abele, & B. Wojciszke (Eds.), *Agency and communion in social psychology. Current issues in social psychology* (pp. 103–116). Routledge.
- Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational Research Review, 27*, 205–217. <https://doi.org/10.1016/j.edurev.2019.03.007>
- Sonnert, G., Fox, M. F., & Adkins, K. (2007). Undergraduate women in science and engineering: Effects of faculty, fields, and institutions over time. *Social Science Quarterly, 88*(5), 1333–1356. <https://doi.org/10.1111/j.1540-6237.2007.00505.x>
- Stadler, M., Herborn, K., Mustafic, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015. An investigation of the validity of the PISA 2015 CPS tasks. *Computers and Education, 157*, Article 103964. <https://doi.org/10.1016/j.compedu.2020.103964>
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581–593. <https://doi.org/10.1177/0956797617741719>
- Suárez-Orozco, C., Suárez-Orozco, M. M., & Todorova, I. (2008). *Learning in a new land: Immigrant students in American society*. Harvard University Press.
- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics, 50*(1), 49–62. <https://doi.org/10.1023/A:1020518104346>
- Van de Werfhorst, H. G., & Mijs, J. B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology, 36*(1), 407–428. <https://doi.org/10.1146/annurev.soc.012809.102538>
- van Hek, M., Buchmann, C., & Kraaykamp, G. (2019). Educational systems and gender differences in reading: A comparative multilevel analysis. *European Sociological Review, 35*(2), 169–186. <https://doi.org/10.1093/esr/jcy054>
- Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis: Research report. *Psychological Science, 18*(1), 19–23. <https://doi.org/10.1111/j.1467-9280.2007.01842.x>
- Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychological Review, 29*(1), 119–140. <https://doi.org/10.1007/s10648-015-9355-x>
- Watkins, D., Cheng, C., Mpofu, E., Olowu, S., Singh-Sengupta, S., & Regmi, M. (2003). Gender differences in self-construal: How generalizable are western findings? *The Journal of Social Psychology, 143*(4), 501–519. <https://doi.org/10.1080/00224540309598459>
- Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 7, pp. 161–198). Elsevier. <https://doi.org/10.1016/bs.adms.2019.05.002>
- Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science, 21*(4), 258–262. <https://doi.org/10.1177/0963721412447622>
- Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 46, pp. 55–123). Academic Press.
- Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology, 41*(1), 331–357. <https://doi.org/10.1146/annurev-soc-071312-145659>
- Yang, K., & Girgus, J. S. (2019). Are women more likely than men to care excessively about maintaining positive social relationships? A meta-analytic review of the gender difference in sociotropy. *Sex Roles, 81*(3–4), 157–172. <https://doi.org/10.1007/s11199-018-0980-y>

Received July 21, 2021

Revision received October 18, 2022

Accepted November 4, 2022 ■