



## Societal level gender inequalities amplify gender gaps in problem solving more than in academic disciplines

Francesca Borgonovi<sup>a,\*</sup>, Samuel Greiff<sup>b</sup>

<sup>a</sup> Institute of Education, University College London, United Kingdom and Directorate for Education and Skills, OECD, France

<sup>b</sup> Cognitive Science and Assessment, ECCS Unit, University of Luxembourg, Luxembourg

### ARTICLE INFO

#### Keywords:

Gender gaps  
Problem solving  
Mathematics  
Cross-national  
Gender inequality  
PISA

### ABSTRACT

Although digital transformations can help narrow existing gender gaps in labour market outcomes, this change depends, among several factors, on the extent to which females have the skills to make the most of new opportunities. An important such skill is problem solving. We examined gender gaps in cognitive and attitudinal dimensions of problem solving and how between-country differences in such gaps are related to societal level gender inequality. This study involved 237,115 students from 42 countries surveyed in the 2012 round of the Programme for International Student Assessment (PISA). Analyses revealed that, on average across the countries considered, males outperformed females on cognitive dimensions of problem solving ( $d = 0.127$ ) and held more positive attitudes towards problem solving ( $d = 0.193$ ). However, gender gaps varied across countries. In countries with greater gender inequality, the gender gap in the problem solving performance of 15-year-olds in favour of males was more pronounced than in countries with lower levels of gender inequality ( $r = 0.27$ ). The association between country-level gender inequality and the gender gap in mathematics, reading, and science was  $r = 0.20$  for mathematics,  $r = 0.22$  for reading, and  $r = 0.02$  for science. Males' advantage in problem-solving performance was in addition to any relationship between gender inequality and the gender gap in text comprehension, mathematics, and the country's level of economic development. By contrast, the gender gap in problem solving attitudes in favour of males was smaller in countries with greater gender inequality ( $r = -0.42$ ).

The computerization and automation of routine tasks are reshaping the skills that are needed to participate successfully in the labour market so that there is now a markedly higher share of nonroutine tasks (Autor, Levy, & Murnane, 2003; Ikenaga & Kambayashi, 2010; Spitz-Oener, 2006). These types of tasks require individuals to react to dynamically changing situations and to engage in fast-paced decision-making. Although digital transformations could empower women and help narrow existing gender gaps in labour market outcomes, this change depends, among several factors, on the extent to which girls and women have the skills and competencies to make the most of new opportunities (Black & Spitz-Oener, 2010; Goldin, 2014; Weinberg, 2000).

Proficiency in problem solving is a skill that appears to be essential for success in modern labour markets (Felstead, Gallie, Green, & Inanc, 2013). Problem solving reflects individuals' capacity to deal with problems that they have not previously encountered, consisting of multiple elements tied by connections that are not immediately transparent and

that can change due to external interventions or the internal dynamics of the problem situation (Greiff et al., 2014; Greeno, 1978; Jonassen, 2011). In problem solving, the use of prior experience alone is not sufficient, and individuals have to engage in active and strategic exploration to be successful.

Strong stereotypes on the masculine nature of mathematics and numeracy skills and on the feminine nature of literacy and reading skills exist, and such stereotypes permeate how students learn, how teachers teach, and how individuals see themselves and perceive others (see OECD, 2015, for a review of the literature). However, no stereotypes on the feminine/masculine nature of problem solving have been defined. Analysing the competence and attitudes that boys and girls display in problem solving and the extent to which they vary across countries is therefore instructive for two key reasons. First, such analyses help to identify how societies and education systems are helping to widen or narrow gender gaps in new domains that are important in modern labour markets. Second, such analyses can help articulate the importance

\* Corresponding author at: Department of Social Science, Institute of Education University College London, 55-59 Gordon Square, London WC1H 0NU, United Kingdom.

E-mail address: [f.borgonovi@ucl.ac.uk](mailto:f.borgonovi@ucl.ac.uk) (F. Borgonovi).

<https://doi.org/10.1016/j.intell.2019.101422>

Received 18 August 2019; Received in revised form 7 December 2019; Accepted 13 December 2019

0160-2896/ © 2019 Elsevier Inc. All rights reserved.

of stereotypes in shaping the relative cognitive development of boys and girls in different societies.

The aim of our paper is threefold. First, we examine gender disparities in problem solving and how they compare with gender gaps in school subjects such as mathematics and text comprehension in which stereotypes are strongly defined. Second, we identify whether the gender gap in problem solving differs across countries and, in particular, whether between-country differences reflect broad societal-level inequalities between men and women. Finally, we consider whether societal-level gender inequality explains between-country differences in the size of gender gaps in problem solving more than it does in the case of achievement in school subjects.

## 1. Past research

Knowledge about gender differences in problem solving is scarce, but available studies have suggested that males<sup>1</sup> outperform females (Hyde, 2005; OECD, 2015; Wüstenberg, Greiff, Molnár, & Funke, 2014). By contrast, gender differences in some of the prerequisites of successful problem solving have been well researched. For instance, males have been found to have, on average, better spatial and navigation abilities than females (Baron-Cohen, 2004; Coutrot et al., 2018; Lawton & Hatcher, 2005; Linn & Petersen, 1985; Reilly & Neumann, 2013), to outperform females in abstract information processing (Halpern & LaMay, 2000), to respond more positively to competitive environments (Niederle & Vesterlund, 2010), to be less risk adverse (Byrnes, Miller, & Schafer, 1999; Fisk, 2018), and to be more proficient in mathematics (see Halpern et al., 2007; Else-Quest, Hyde, & Linn, 2010; Ceci, Ginther, Kahn, & Williams, 2014, for comprehensive reviews).

Several empirical and theoretical studies in economics, psychology, and sociology have investigated the gender achievement gap in mathematics (Ceci et al., 2014; Halpern, 2012; Halpern et al., 2007; Lindberg, Hyde, Petersen, & Linn, 2010; Machin & Pekkarinen, 2008; Miller & Halpern, 2014). Gender differences in favour of males in average mathematics achievement and general achievement among individuals with very high levels of ability decreased during the 1970s and 1980s (Lakin, 2013; Wai, Cacchio, Putallaz, & Makel, 2010). Findings indicate that the gender gap in average mathematics performance is now small or nonexistent in many countries (Miller & Halpern, 2014) although young males continue to outnumber young females among high achievers (OECD, 2015; Wai et al., 2010).

Because the literature on gender gaps in mathematics is large and there is a strong link between some of the cognitive processes that form the basis of mathematical reasoning and proficiency in problem solving, such literature can importantly inform the research on gender gaps in problem solving. Contrary to mathematics, a subject for which strong stereotypes exist on the relative abilities of males and females, no stereotypes exist with respect to domain-general problem solving.

Some researchers have highlighted biological differences between males and females as a source of gender gaps in mathematics and other cognitive abilities such as spatial aptitude. For example, it has been suggested that performance in spatial tasks is influenced by androgen levels in utero and may therefore give males an advantage in domains that rely on such cognitive abilities (see Auyeung, Lombardo, & Baron-Cohen, 2013; Valla & Ceci, 2011; Vuoksimaa et al., 2010). However, females who are exposed to androgen levels that are comparable to those experienced by males do not enjoy similar performance advantages, and gender gaps vary greatly both over time and across cultures, suggesting that biological factors play a limited role (Hines et al., 2003; Vuoksimaa et al., 2010).

<sup>1</sup> In this section, the ages of participants varied, so we decided to use the term males for conciseness rather than specifically identifying boys, male adolescents, and men. Similarly, we used the term females rather than specifically identifying girls, female adolescents, and women.

Although gender gaps in early spatial abilities and mathematics are well-documented (Ceci et al., 2014), there is evidence that gender gaps can work to the advantage of males when problems are framed as geometry problems and to the advantage of females when the same tasks are framed as art tasks (Huguet & Regner, 2009). In fact, the literature highlights large variations in the strength of gender gaps in a broad array of cognitive skills depending on how skills are measured, how problems are framed, and the sociocultural experiences children are exposed to (Ceci et al., 2014; Coutrot et al., 2018; Eagly & Wood, 1999; Else-Quest et al., 2010; Halpern, 2012; Hedges & Nowell, 1995; Miller & Halpern, 2014; Shih, Pittinsky, & Ambady, 1999).

Furthermore, it has been suggested that females display different levels of performance on mathematics tests depending on the extent to which their gender is made salient (i.e., lower performance is associated with the salience of gender; Steele, 1997). Stereotype threat is a phenomenon by which the members of a social group have an awareness that others expect them to do poorly, and this awareness then leads them to perform below their ability even when they themselves do not endorse such negative expectancies (Spencer, Steele, & Quinn, 1999; Steele, 1997). Meta-analytic evidence has suggested that the effect of stereotype threat on the performance of girls in mathematics is small ( $d = 0.22$ ; Flore & Wicherts, 2015). Stereotypes arise from the observation that objective differences in outcomes exist across groups (Jussim, 2012), but these observations are often turned into culturally and socially transmitted deterministic beliefs that shape expectations and action (Johnson & Wilson, 2019) given specific sociocultural contexts.

Among the sociocultural factors that have been analysed is societal-level gender inequality. Previous research has shown that gender gaps in favour of boys on standardised math tests are larger in countries where women have lower participation in the labour market and are less represented in political institutions (Breda, Jouini, & Napp, 2018; Else-Quest et al., 2010; Guiso, Monter, Sapienza, & Zingales, 2008), although a number of studies have indicated no association between societal-level gender inequality and gender gaps in mathematics achievement (Meisenberg, 2016; Stoet & Geary, 2015). Conceptually, the correlation between societal-level gender inequality and the gender gap in mathematics achievement has been traced to stereotypes of men's and women's roles that, in turn, are reflected in societies in which women participate in the labour market at different levels and where women are less represented in political institutions than men.

According to expectancy value theory, the motivation individuals exert when engaging in a task or pursuing an occupation or field of study is a function of their expected performance and the degree to which they value the task (Eccles & Wigfield, 2002). Stereotypes operate at several levels to lower the performance expectancies of males and females in subjects and fields that are counterstereotypical and to lower the value that males and females assign to excelling in fields that are counterstereotypical. First, they reduce incentives for females to invest time and effort in counterstereotypical skills because they will not have adequate returns on such investments in the labour market, thus lowering the value assigned by females to such fields. This is because finding employment in sectors requiring such skills will be more difficult for females because many prospective employers will not expect females to excel in their sectors. Second, in the presence of strong stereotypes, females will be socialised into not valuing mathematics and activities associated with quantitative abilities. Third, stereotypes can lead females to hold lower expectations of their ability to acquire skills that do not conform to stereotypical notions of feminine skills—again reducing their incentives to (and thus they fail to) invest in building such skills. Gender differences in subjective orientations towards mathematics and science are already present at age 5 (Perez-Felkner, McDonald, Schneider, & Grogan, 2012), and by age 15, male students are twice as likely to expect to work in mathematics and engineering careers than female students, although gender gaps in career orientations differ greatly across countries and over time (OECD, 2016).

Fourth, in the presence of strong stereotypes, females might also not receive the support they need to excel in mathematics from their teachers and families (Else-Quest et al., 2010), thus lowering the extent to which their abilities come to fruition given their endowed potential. Fifth, societal-level gender inequality may also influence gender gaps in achievement through stereotype threat. In societies with higher levels of gender inequality, stereotypes on gender differences in aptitudes may be more prevalent (Correll, 2001).

An apparent paradox is that although the literature has revealed large gender gaps in mathematics achievements, gender gaps in mathematics attitudes and self-beliefs are largest in the presence of greater gender equality (Else-Quest et al., 2010; Stoet, Bailey, Moore, & Geary, 2016), possibly because of differences in social comparisons for females in such contexts (Guimond et al., 2007). The finding that gender gaps in mathematics attitudes and self-beliefs are strongest in more gender-equal countries mirrors findings on the extent to which gender differences in preferences for risk, degree of competitiveness, trust, altruism, patience, and reciprocity are related to societal-level gender equality (Falk & Hermle, 2018). This finding has been explained by the suggestion that a more egalitarian distribution of material and social resources might enable women and men to express gender-specific preferences.

## 2. The present study

Because of lack of cross-country comparable data on general problem-solving abilities, so far it has not been possible to test whether societal inequalities influence females' acquisition of general cognitive skills (rather than just mathematics or other curricular domains such as text comprehension). It has similarly not been possible to determine whether the root of gender disparities in skills lies in broader social influences rather than in the education system.

The focus on mathematics achievement is narrow and mostly driven by the availability of internationally comparable data on math achievement among school-aged males and females. Key explanations for the reason why societal-level gender inequalities shape the emergence and development of gender gaps in achievement rely on the roles that societal norms and attitudes play in shaping and reinforcing gender stereotypes and, by doing so, the opportunities males and females have to develop their skills and the incentives they have to do so. The implication of the focus on mathematics is that societal-level disparities impede females' abilities in mathematics but not necessarily in other domains and may in fact result in greater segregation and 'specialisation' by gender with females doing better in fields that fit stereotypically feminine ideals.

But what if societal-level gender inequalities also hold females back in more general cognitive skills such as problem solving for which strong stereotypes do not exist and for which opportunities and incentives are not yet well-defined? Our work extends recent evidence that gender gaps in navigation abilities are wider in countries with societal-level gender inequalities (Coutrot et al., 2018) and that gender differences in attitudes and preferences depend on societal-level gender inequalities (Falk & Hermle, 2018).

We used data from a large-scale international assessment of 15-year-old students, the Programme for International Student Assessment (PISA), which was administered in 42 countries. We used PISA to identify the gender gap in problem solving, the extent to which the size of the gender gap varies across countries, whether such variation is systematically related to the level of broad societal-level inequality, and how gender gaps in problem solving (and their association with societal-level gender inequalities) compare with gender gaps in academic domains such as mathematics.

Newell and Simon (1972) defined the process of problem solving as: 'a given state is transformed into a goal state by applying a sequence of consciously selected actions'. According to Mayer and Wittrock (2006), problem solving is a cognitive process that occurs internally within a

person's information-processing system. Therefore, information-processing resources (e.g., attention or the capacity of the human mind to store and manipulate different types of information simultaneously) are important for successful problem solving (Wiley & Jarosz, 2012). Likewise, problem solving is a process that is based on representing and manipulating knowledge structures such that problem solving depends on what people know about a particular problem and can use to solving it (Goode & Beckmann, 2010).

In a comprehensive review of problem solving Greiff and colleagues (2017) indicate how several dimensions differentiate problems and problem-solving processes (VanLehn, 1989). First, problems can be knowledge-lean or knowledge-rich. In knowledge-lean problems, all information that is needed is already included in the description of the problem. Knowledge-rich problems require the problem solver to apply background knowledge or to search for (additional) information that can be used to understand and to solve the problem. Second, some problems can be solved with a single action, whereas others require multiple steps to reach a solution. Third, some problems are static in the sense that the problem state changes only in response to the problem solver's actions, whereas in dynamic problems, the problem state comprises many factors that change independently of the problem solver's actions (Funke, 2010). Finally, problems can be well-defined such that there is a well-specified given state, goal state, and set of operators that can be applied to move from the given information to the goal. Alternatively, problems can be ill-defined, in which case these aspects are unknown or ambiguous (Greeno, 1978; Jonassen, 2011). These classifications of problems need to be considered when defining and, in a second step, assessing problem-solving skills.

PISA's problem-solving assessment, which we used in our analyses, reflects the features of real-world problems that individuals face in their daily lives: They are knowledge-rich, involve multiple steps, and can be both static and dynamic as well as more or less well-defined. More specifically, thanks to the computer-based delivery, around two thirds of the assessment tasks are dynamic, and one third are static.

At the conceptual level, problem solving is defined in PISA as the 'capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential' (OECD, 2014a, p. 30). Thus, there are cognitive and attitudinal dimensions to problem solving. Examples of tasks included in the computer-based test used to derive the cognitive dimension of problem solving are available in the OECD (2014a). Whereas the problem-solving assessments were strongly associated with the other cognitive domains assessed in PISA (among countries in our sample  $r = 0.81$  between problem solving and science;  $r = 0.84$  between problem solving and mathematics, and  $r = 0.77$  between problem solving and reading), these associations were less pronounced than was the case for the other assessment domains ( $r = 0.91$  between science and mathematics;  $r = 0.89$  between science and reading, and  $r = 0.87$  between reading and mathematics). Furthermore, around 32% of the variation in problem solving reflects skills that are uniquely captured by the problem-solving assessment, whereas 68% of the individual variation in problem-solving achievement is shared with the other three assessment domains (OECD, 2014a). These results indicate that whereas the PISA problem-solving assessment may measure a unique set of abilities, it is also highly correlated with other measures of achievement in PISA, thus providing further empirical evidence in support of the homogeneity of cognitive abilities (g-factor; Jensen, 1998; Rindermann, 2007).

To measure attitudes towards problem solving, we used information that PISA participants provided in a background questionnaire. Information was collected on whether a participant assessed him- or herself as someone who can handle a lot of information, is quick to understand things, seeks explanations for things, can easily link facts together, and likes to solve complex problems. Responses consisted of: This person is 'very much like me', 'mostly like me', 'somewhat like me',

'not much like me', or 'not at all like me'. We used self-reports to develop an indicator of openness to problem solving based on self-reports.

### 3. Method

#### 3.1. Data

We used data from the 2012 cycle of PISA, a large-scale, cross-national assessment of the performance of 15-year-old students. In the 2012 cycle, on top of the core paper-based assessment instruments in mathematics, reading, and science, students in 42 countries were administered a computer-based test in problem solving, and 31 of these countries also administered a computer-based test in reading and mathematics. The problem-solving test was designed to require minimal literacy, numerical, and mathematics abilities.

PISA samples are representatively drawn from the population of 15-year-old students in each country. The total sample size in the 42 countries that participated in both the core and the PISA problem-solving assessment was 237,115 students. Sample size was 183,151 in the 31 countries that administered the computer-based reading and mathematics assessment. Within countries, a two-stage sampling procedure selected the participants who took the PISA test. Weighted samples in each country are representative of students who are enrolled in the seventh grade or above and are between 15 years 3 months and 16 years 2 months at the time of the assessment. All variables and cases used in our analyses were derived from the public-use files for the PISA 2012 survey, which can be downloaded from: <http://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm>

Table 1 presents descriptive statistics for each country included in our analyses. Overall, participants included in our main specification (sample of students in the 42 countries who participated in the problem-solving assessment) consisted of 49.87% female students and 50.13% male students. We excluded 34,208 students who were either foreign born and/or had foreign-born parents because of evidence that these students may be influenced by the social norms of their country of origin as well as their destination country (Nolleberger, Rodríguez-Planas, & Sevilla, 2016) and, thus, might not be well-suited to answer the research questions at hand.

#### 3.2. Materials

Students in all PISA participating countries took a 2-hr paper-based test containing a range of tasks in reading, mathematics, and science and completed a background questionnaire. The PISA assessment tasks were developed by an international consortium in collaboration with multidisciplinary expert groups and under the supervision of the PISA governing board, a body of the OECD composed of representatives of participating countries.

In countries participating in the problem-solving assessment, students who were selected to take the computer-based test were administered an additional 40-min test containing a range of problem-solving tasks in order to cover the cognitive dimension of problem solving. They also answered a range of reading and math questions that were different from the questions administered in the 2-hr paper-based assessment.

Students' openness to problem solving was evaluated via self-reports. These were part of the main PISA student background questionnaire, a 30-min paper-and-pencil questionnaire administered after the core assessment. The background questionnaire consisted of a set of socioeconomic and demographic questions administered to all students and three sets of additional questions. Each student randomly received a questionnaire containing two out of the three sets of additional questions. The rotation was organised to increase the coverage of information gathered at the population level without increasing the burden for individual respondents. The randomisation process worked as intended, and no differences in background characteristics could be

detected between students assigned to any of the three forms (OECD, 2014b).

#### 3.3. Variable descriptions

##### 3.3.1. Dependent variables: problem-solving performance and openness to problem solving

Students' problem-solving performance scores were derived from a one-parameter IRT model that mapped student responses to their underlying unobserved ability (cognitive dimension). Students' patterns of responses to specific problem-solving tasks were used to generate plausible value scores of students' problem-solving performance (OECD, 2014b). Problem-solving performance in PISA is standardised to have a mean of 500 and a standard deviation of 100 across OECD countries. In all our analyses, we report standardised effect sizes by dividing PISA achievement scores by the pooled standard deviation in our sample ( $SD = 102$ ). This means that whenever we report gender differences, we effectively report Cohen's  $d$  statistics. In all models, gender gaps represent the difference between males and females. Therefore, the difference between our estimates and Cohen's  $d$  is that we interpret estimates with a positive sign to indicate an advantage for males<sup>2</sup> and a negative sign to indicate an advantage for females.

The openness to problem-solving index (attitudinal dimension) was derived using a one-parameter IRT model and was standardised to have a mean of 0 and  $SD$  of 1 across OECD countries. The index has a good degree of reliability (Cronbach's alpha higher than 0.80 in most countries; for details on country-specific reliability indices, see Table A.1). Because the index had a standard deviation of 1, gender differences correspond to Cohen's  $d$  statistics. In all models, gender gaps represent the difference between male and females (positive  $d$  estimates indicate an advantage for males and negative  $d$  estimates indicate an advantage for females).

##### 3.3.2. Key explanatory variables: gender and country-level gender inequality

Students reported their gender, and the information was verified with sampling forms. Country-level gender inequality was measured with the Gender Inequality Index (GII), a composite index characterising levels of gender inequality across countries. The GII index was developed by the United Nations Development Programme (UNDP). The GII is based on three dimensions: female reproductive health, measured by the maternal mortality ratio and adolescent birth rates; gender empowerment, measured by the relative proportion of parliamentary seats occupied by women and men and the relative proportion of women and men 25 years of age and older with at least some secondary education; and gender labour market status, measured by the relative labour force participation rate of female and male populations 15 years of age and older (UNDP, 2016). The GII ranges from 0 to 1, where higher values indicate greater gender inequality. To verify the robustness of the findings, the main results, which were based on the composite GII, are accompanied by analyses calculated with the original indicators for each of the three different components. Both the GII and its components were standardised to have a mean of 0 and an  $SD$  of 1 so that estimated coefficients could be interpreted in terms of changes in  $SD$ s.

#### 3.4. Control variables

In the background questionnaire, students were asked for information on their socioeconomic status. PISA derives an aggregate indicator

<sup>2</sup> When referring to the participants of the present study, we decided to use the term males for conciseness rather than male adolescents or male students. Similarly, we used the term females rather than female adolescents or female students.



**Table 1**  
Country-specific descriptive statistics.

Country code	Country name	N	% boys	Mean SES	Mean SES (females)	Mean problem solving performance	Mean openness to problem solving	Mean computer-based math performance	Mean computer-based reading performance	Mean paper-based math performance	Mean paper-based reading performance
ARE	United Arab Emirates	5523	49%	0.275	0.246	374	0.329	405	365	398	407
AUS	Australia	11,948	52%	0.260	0.252	521	-0.102	503	516	498	507
AUT	Austria	4015	50%	0.186	0.172	515	0.049	515	490	515	498
BEL	Belgium	7368	50%	0.222	0.199	519	-0.307	520	513	525	518
BGR	Bulgaria	5258	52%	-0.276	-0.264	402	0.371	N/A	N/A	439	437
BRA	Brazil	5469	48%	-1.153	-1.227	425	0.234	419	431	394	410
CAN	Canada	17,758	50%	0.451	0.445	528	0.114	520	531	518	521
CHL	Chile	6783	48%	-0.582	-0.615	448	0.174	432	452	423	441
COL	Colombia	9042	47%	-1.263	-1.329	399	0.184	397	396	377	404
CZE	Czech Republic	5134	51%	-0.063	-0.055	510	-0.194	N/A	N/A	500	494
DEU	Germany	4460	51%	0.282	0.287	513	0.172	513	497	519	512
DNK	Denmark	5670	50%	0.490	0.485	503	0.012	502	500	506	501
ESP	Spain	9265	50%	-0.131	-0.130	481	0.037	481	472	489	491
EST	Estonia	4398	49%	0.113	0.088	518	0.047	519	526	523	519
FIN	Finland	7559	51%	0.386	0.406	526	-0.109	N/A	N/A	522	527
FRA	France	3960	49%	0.046	0.036	521	-0.192	516	518	504	515
GBR	United Kingdom	3633	50%	0.304	0.280	521	-0.047	N/A	N/A	497	501
HKG	Hong Kong-China	3077	55%	-0.567	-0.634	542	-0.253	552	551	563	544
HRV	Croatia	4417	51%	-0.304	-0.333	467	-0.024	N/A	N/A	473	487
HUN	Hungary	4733	48%	-0.259	-0.290	459	0.170	470	450	476	488
IRL	Ireland	4524	51%	0.120	0.108	499	-0.035	493	521	502	524
ISR	Israel	4191	49%	0.201	0.141	450	0.356	445	458	465	484
ITA	Italy	5098	54%	0.007	-0.002	513	-0.059	502	507	492	492
JPN	Japan	6328	53%	-0.071	-0.072	552	-0.729	539	545	537	538
KOR	Korea	5031	53%	0.012	-0.025	561	-0.371	553	555	554	536
MAC	Macao	1908	51%	-0.691	-0.627	536	-0.404	533	506	526	494
MNE	Montenegro	4467	50%	-0.255	-0.297	406	0.618	N/A	N/A	408	422
MYS	Malaysia	5111	48%	-0.714	-0.725	423	-0.202	N/A	N/A	421	398
NLD	Netherlands	3986	51%	0.298	0.289	518	-0.102	N/A	N/A	529	517
NOR	Norway	4233	51%	0.519	0.491	508	0.181	501	506	494	508
POL	Poland	4597	49%	-0.211	-0.223	481	0.363	489	477	517	518
PRT	Portugal	5323	51%	-0.473	-0.495	496	0.157	491	489	490	490
RUS	Russian Federation	4662	49%	-0.098	-0.107	490	0.049	491	466	484	478
SGP	Singapore	4570	51%	-0.340	-0.349	560	-0.033	562	567	569	539
SRB	Serbia	4302	50%	-0.296	-0.330	473	0.464	N/A	N/A	448	444
SVK	Slovakia	4644	52%	-0.184	-0.198	483	-0.321	497	474	482	463
SVN	Slovenia	5399	52%	0.121	0.130	480	0.079	490	475	505	485
SWE	Sweden	4054	51%	0.357	0.348	499	0.104	496	506	486	492
TAP	China	6016	49%	-0.397	-0.387	534	-0.334	538	520	560	523
TUR	Turkey	4806	51%	-1.459	-1.456	454	0.209	N/A	N/A	448	476
URY	Uruguay	5291	47%	-0.885	-0.954	403	0.040	N/A	N/A	409	411
USA	United States	3976	51%	0.327	0.331	510	0.200	501	514	484	498

N/A: No data available because the computer-based test of mathematics and reading was not administered.

of socioeconomic status (SES) using information provided by students on parental educational attainment, occupational status, and household resources. The SES indicator is standardised to have a mean of 0 and *SD* of 1 across OECD countries. The SES values for the females and males in our sample were -0.304 and -0.244, respectively (the values were negative because our sample was not limited to OECD countries). Country-specific summary statistics are reported in Table A.2 in the Appendix.

### 3.5. Statistical analyses

We followed standard procedures to analyse the PISA data, taking into account plausible values and sampling (OECD, 2014b). Our key dependent variables were students' problem-solving performance (cognitive dimension) and their reported openness to problem solving (attitudinal dimension).

For each domain, achievement scores were represented in PISA by a set of five plausible values, which are essentially imputed values of

proficiency (OECD, 2014b). This means that in each model in which we estimated gender gaps in achievement, we fit five sets of models, one for each plausible value, and then we combined these values using Rubin's rule as per OECD recommendations (Little & Rubin, 1987; OECD, 2014b). In all models, whether we estimated achievement or attitudes towards problem solving, we calculated robust standard errors to account for the clustering effect of the two-stage sampling in PISA, calculating standard errors using balanced repeated replication (BRR) and students' final weights. This allowed us to correctly account for the nested structure of the PISA data and the sampling frame used in each participating country (OECD, 2014b). Because we focused on the pooled PISA sample, we used rescaled probability weights (Rutkowski, Gonzalez, Joncas, & von Davier, 2010) so that we could equally weight the extent to which each country contributed to the pooled model. Because we were interested in examining the contribution of country-level factors to the width of gender gaps in problem-solving proficiency and attitudes, and the sampling design in PISA involves sampling students from schools in a specific country, we modeled interactions

between the gender variable (individual-level predictor) and indices of gender inequality (country-level predictor) by considering clustering via BRR. Our approach resulted in standard errors that were less biased than standard errors estimated with a three-level multilevel model (we also fit three-level hierarchical models, and these results can be requested from the authors).

We present results on gender differences in problem solving both on a bivariate level and after controlling for differences in mathematics and reading performance. We did this by introducing mathematics and reading performance as covariates in regressions where problem-solving performance and problem-solving attitudes were the dependent variables (in all regression models, we matched the first plausible value in problem solving with the first plausible value in mathematics and reading). These analyses allowed us to identify gender differences that were specific to problem-solving performance and attitudes and were not a simple reflection of gender differences in other information processing abilities. The results, which can be requested from the authors, were virtually identical to estimates obtained from an alternative specification in which we accounted for the potential influence of curricular subjects by generating a variable that represented the difference between problem-solving performance and each of the curricular subjects). Furthermore, by examining gender gaps in problem-solving performance and attitudes after controlling for mathematics and reading, we were able to identify the unique association between gender inequality at a country level and gender gaps in problem solving after accounting for any effect gender inequality may have on gender gaps in mathematics and reading. We present additional results in which we controlled for the socioeconomic status of the students and the schools they attended.

In a subsample of countries (31 out of 42), on top of the core paper-based assessments and the computer-based problem-solving assessment, students also took a computer-based assessment of reading and mathematics. Although the subsample of countries with available data on computer-based reading and mathematics was not random because it was a function of countries' decision to participate in such additional tests, the decision was primarily driven by cost considerations and interest in having results in such additional domains, given the expected broad transition of all PISA instruments to computer delivery in 2015. As a robustness check, for those students who took the test in countries where the computer-based assessment included computer-based reading and mathematics, when estimating our models, we controlled for students' performance on the computer-based reading and mathematics tests to ensure that our findings did not reflect gender differences in preferences for computers or in the mode of assessment.

#### 4. Results

Mean comparisons of performance in problem solving across gender revealed that males outperformed females in 19 out of the 42 countries examined, whereas females outperformed males in 4 countries ( $p \leq .05$ ). Results that are pooled across all countries with available

data are reported in Table 2, whereas country-specific estimates can be found in Tables A.3 and A.4 in the Appendix. The Cohen's  $d$  estimate for gender obtained while not controlling for individual and school-level factors across the 42 countries with available data was  $d = 0.127$ , indicating an advantage for males. By comparison, the gender gap in math across the same 42 countries was  $d = 0.126$ , and the value was  $d = -0.310$  (in favour of females) in reading. The residual gender gap in problem-solving performance, obtained after accounting for student performance in the paper-based math and reading test was  $d = 0.123$ . Accounting for student performance in the computer-based math and reading tests resulted in a gender gap in problem solving of  $d = 0.146$  (results apply to the sample of 31 countries that administered the computer-based assessment). Additionally controlling for individual- and school-level SES did not affect the estimates ( $d = 0.126$ ). Table 1 shows that the results were consistent and were not very affected by the introduction of covariates and by the estimation method we applied.

Our results revealed a gender gap not only in problem-solving performance but also in attitudes towards problem solving. In fact, gender differences appeared to be particularly large in 15-year-old students' reports of their openness to problem solving. Males reported greater openness to problem solving than females in 35 of the 42 countries ( $p \leq .05$ ), and in no country did females report greater openness to problem solving than males. The gender gap in the openness to problem-solving index was  $d = 0.193$ , indicating higher levels for males. Table 2 shows that gender differences in attitudes towards problem solving occurred in addition to any differences between males and females in problem-solving achievement: The gender gap in problem solving attitudes remained stable when we controlled for problem-solving performance ( $d = 0.176$ ), problem solving, mathematics, and reading performance ( $d = 0.181$ ), and when we controlled for performance in the three domains as well as for individual- and school-level SES ( $d = 0.169$ ).

Although on average males appeared to show better problem-solving performance and more positive attitudes towards problem solving than females did, Fig. 1 reveals heterogeneity in gender gaps across countries with regard to both the cognitive and attitudinal dimensions. The distribution of the gender gap in problem-solving performance (Panel a) was similar to what was observed for mathematics (Panel c), whereas the distribution shifted towards the right and was less dispersed for problem-solving attitudes (Panel b).

Table 3 illustrates country-level associations between gender inequality and the gender gap in problem-solving performance, problem-solving attitudes, as well as the gender gap in the curricular domains of mathematics, reading, and science. Results indicate that at the country level, the association between gender inequality and the gender gap in problem-solving performance was  $r = 0.27$ , stronger than the association between gender inequality and the gender gap observed for mathematics ( $r = 0.20$ ), reading ( $r = 0.22$ ), and science ( $r = 0.03$ ). The country-level association between gender inequality and the gender gap in problem-solving attitudes was strong and negative ( $r = -0.42$ ). This means that in countries with greater gender equality, the gender gap in

**Table 2**

The gender gap in problem-solving performance and attitudes.

	Problem-solving performance ( <i>N</i> countries)		Attitudes to problem solving ( <i>N</i> countries)
Gender gap (M-F)	0.127 (42)	Gender gap (M-F)	0.193 (42)
Gender gap (M-F) controlling for paper-based math and reading	0.123 (42)	Gender gap (M-F) controlling for PS performance	0.176 (42)
Gender gap (M-F) controlling for paper-based math and reading, and individual and school SES	0.126 (42)	Gender gap (M-F) controlling for PS performance, paper- based math and reading	0.181 (42)
Gender gap (M-F) controlling for computer-based math and reading	0.146 (31)	Gender gap (M-F) controlling for PS performance, paper- based math and reading, and individual and school SES	0.169 (42)
		Gender gap (M-F) controlling for PS performance, computer-based math and reading	0.163 (31)

Source: PISA 2012 database. M = males; F = females.



**Table 3** Country-level associations between gender inequality and gender gaps in problem solving, reading, mathematics, and science.

	Gender Inequality Index	Gender gap in problem-solving performance	Gender gap in mathematics performance	Gender gap in reading performance	Gender gap in science performance	Gender gap in problem-solving attitudes
Gender Inequality Index	1					
Gender gap in problem-solving performance	0.267	1				
Gender gap in mathematics performance	0.196	0.712	1			
Gender gap in reading performance	0.222	0.689	0.701	1		
Gender gap in science performance	0.026	0.761	0.801	0.852	1	
Gender gap in problem-solving attitudes	-0.418	-0.004	0.221	0.238	0.305	1

Note. Country-level correlations. The gender gap refers to males-females such that positive numbers on the gender gap indicate an advantage for males and negative numbers an advantage for females. The Gender Inequality Index increases as the difference in outcomes between males and females increases.

Geary, 2015). When problem-solving performance was controlled for, very different results emerged: Greater country-level gender inequality was associated with a smaller gender gap in math performance in favour of males, rather than a larger gap.

Table 4 also indicates that societal-level gender inequality was associated with overall lower levels of achievement in mathematics irrespective of the set of controls that was introduced.

When controlling for the full set of individual, school, and country-level controls, our estimates indicate that the gender gap in problem solving was smallest in the Netherlands ( $d = 0.079$ ; the least gender unequal country in our sample based on GII) and widest in Colombia ( $d = 0.150$ ; the most gender unequal country). Results were robust to the specific indicator of gender inequality examined.

Table 4 highlights that males had more positive attitudes towards problem solving than females and that the estimated gender gap in students' openness to problem solving was robust to the inclusion of individual-level controls for ability on the PISA problem-solving, mathematics, and reading test. Contrary to the findings estimated for problem-solving performance, the gender gap in problem-solving attitudes appeared to be smaller in countries with greater gender inequality and larger in countries with greater gender equality. These results were similar when different measures of gender inequality were used and no matter whether individual performance and country-level GDP were controlled for or not. For example, the positive main effect of GII and the negative cross-level interaction (between GII and the male identifier) imply that greater gender inequality was associated with overall more positive attitudes towards problem solving, in particular among females.

### 5. Discussion

We found that gender gaps in cognitive aspects of problem solving tended to be larger in societies with lower levels of gender equality but that gender gaps in attitudinal dimensions of problem solving tended to be smaller in societies with lower levels of gender equality. This finding mirrors previous estimates in which country-level gender inequality measures were related to performance in mathematics, attitudes towards mathematics, and mathematics self-beliefs (Else-Quest et al., 2010; Stoet et al., 2016) as well as gender differences in the willingness to take risks, patience, altruism, reciprocity, and trust (Falk & Hermle, 2018). We also found that country-level gender inequality was associated with gender disparities in problem-solving performance on top of and beyond any effect country-level gender inequality has on curricular subjects such as mathematics and text comprehension. In fact, the gender gap in curricular subjects such as mathematics tends to be narrower in countries that are more gender unequal, once differences in problem-solving performance are controlled for.

Problem solving lacks a disciplinary or curricular nature, there are no stereotypes that identify problem solving as an area in which either males or females excel, and problem solving plays no role as a gate-keeper. Therefore, the mechanisms through which gender inequalities shape the gender gap in problem-solving ability and attitudes must differ from those identified for mathematics (Else-Quest et al., 2010) and are not related to occupational aspirations, role modelling, or performance expectancies. Furthermore, the fact that gender gaps in problem-solving performance are more strongly associated with societal-level gender inequality than gender gaps in performance in curricular subjects suggests that the school environment and instruction may even play a narrowing role with regard to gender gaps in curricular abilities. This finding stands in marked contrast with hypotheses put forward in the literature on the roles played by course choices, teaching practices, and peer effects in education as determinants of gender gaps and, in turn, on how all these factors are influenced by societal gender norms.

A possible reason that could explain our findings lies in the role social structures have in organising play and children's interactions



**Table 4**  
Gender inequality and the gender gap in problem-solving performance and attitudes.

	Controls for	Male		GII		Male*GII interaction	
		b	(SE)	b	(SE)	b	(SE)
Panel A. Dependent variable: Problem solving performance. $N = 218,493$							
Model A1	Individual and school	0.071***	(0.017)	-0.103***	(0.015)	0.199**	(0.073)
Model A2	Individual, school, math & reading	0.092***	(0.014)	0.040**	(0.012)	0.130*	(0.052)
Model A3	Individual, school, GDP	0.072***	(0.017)	-0.103***	(0.015)	0.245***	(0.073)
Model A4	Individual, school, GDP, math and reading	0.082***	(0.013)	0.059***	(0.013)	0.172***	(0.052)
Panel B. Dependent variable: Problem solving attitudes. $N = 142,607$							
Model B1	Individual and school	0.285***	(0.020)	0.237***	(0.008)	-0.362***	(0.083)
Model B2	Individual, school, problem solving performance	0.268***	(0.020)	0.253***	(0.007)	-0.386***	(0.084)
Model B3	Individual, school, problem solving, math & reading	0.250***	(0.023)	0.287***	(0.008)	-0.362***	(0.084)
Model B4	Individual, school, GDP	0.282***	(0.020)	0.229***	(0.009)	-0.355***	(0.083)
Model B5	Individual, school, GDP, problem solving performance	0.265***	(0.020)	0.245***	(0.008)	-0.379***	(0.084)
Model B6	Individual, school, GDP, problem solving, math and reading	0.237***	(0.023)	0.290***	(0.009)	-0.356***	(0.085)
Panel C. Dependent variable: Mathematics performance. $N = 218,493$							
Model C1	Individual and school	0.099***	(0.015)	-0.190***	(0.011)	0.068	(0.059)
Model C2	Individual, school, problem solving performance	0.046***	(0.012)	-0.114***	(0.011)	-0.079	(0.044)
Model C3	Individual, school, problem solving & reading	0.274***	(0.009)	-0.089**	(0.008)	-0.066*	(0.032)
Model C4	Individual, school, GDP	0.104***	(0.015)	-0.211***	(0.012)	0.066	(0.058)
Model C5	Individual, school, GDP, problem solving performance	0.052***	(0.011)	-0.136***	(0.011)	-0.077	(0.044)
Model C6	Individual, school, GDP, problem solving and reading	0.028***	(0.003)	-0.105***	(0.007)	-0.066*	(0.031)

Note. Source: PISA 2012 database. Pooled models. Results for individual components of the GII index are available in Tables A.5 and A.6 in the Appendix. Each panel presents the following key results: the standardised gender gap (expressed in terms of the difference in problem-solving performance among males with females being the baseline), the change in problem-solving performance associated with a 1-unit change in the Gender Inequality Index and additional change in problem-solving performance associated with a 1-unit change in the Gender Inequality Index among males. Panel A presents results for problem-solving performance, Panel B presents results for problem-solving attitudes, and Panel C presents results for mathematics performance. Within each panel, each row represents a different model. Each model differs because of the controls that were introduced and that are highlighted in the controls column.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

with peers and adult figures, which can foster (or hinder) the development of important prerequisites of problem solving. Because engagement in sex-specific behaviour has been shown to be context-dependent (Iervolino, Hines, Golombok, Rust, & Plomin, 2005), in countries with low levels of gender equality, boys and girls may be exposed to strong social norms on the suitability of engaging in different play and social activities that match sex-specific stereotypes.

We believe that because societal gender inequality is less predictive of the size of the gender gap in mathematics than in noncurricular domains such as problem solving, schools should not be considered the primary institutions that contribute to gender inequalities through gender differences in course choices and transitions. Rather, our results suggest that education systems may act in ways that somewhat reduce the influence of societal gender inequalities on skill development and acquisition.

## Acknowledgments

Francesca Borgonovi acknowledges support from the British Academy's Global Professorship programme. The opinions expressed by the authors do not represent the views of the Organisation for Economic Cooperation and Development.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.intell.2019.101422>.

## References

- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*, 1278–1333.
- Auyeung, B., Lombardo, M. V., & Baron-Cohen, S. (2013). Prenatal and postnatal

- hormone effects on the human brain and cognition. *European Journal of Physiology*, *465*, 557–571.
- Baron-Cohen, S. (2004). *The essential difference*. London: Penguin Books.
- Black, S. E., & Spitz-Oener, A. (2010). Explaining women's success: Technological change and the skill content of women's work. *The Review of Economics and Statistics*, *92*, 187–194.
- Breda, T., Jouini, E., & Napp, C. (2018). Societal inequalities amplify gender gaps in math. *Science*, *359*, 1219–1220.
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*, 367–383.
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, *15*, 75–141.
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *106*, 1691–1730.
- Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., ... Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology*, *28*, 1–6.
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*, 408–423.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127.
- Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412).
- Felstead, A., Gallie, D., Green, F., & Inanc, H. (2013). *Skills at work in Britain: First findings from the Skills and Employment Survey 2012*. London: Centre for Learning and Life Chances in Knowledge Economies and Societies, Institute of Education.
- Fisk, S. R. (2018). Who's on top? Gender differences in risk-taking produce unequal outcomes for high-ability women and men. *Social Psychology Quarterly*, *81*, 185–206.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*, 25–44.
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, *11*, 133–142.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*, 1091–1119.
- Goode, N., & Beckmann, J. F. (2010). You need to know: There is a causal relationship between structural knowledge and control performance in complex problem solving tasks. *Intelligence*, *38*, 345–352.
- Greeno, J. (1978). Natures of problem-solving abilities. In W. Estes (Ed.), *Handbook of learning and cognitive processes*. Hillsdale, NJ: Erlbaum.
- Greiff, S., Wüstenberg, S., Csapo, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21<sup>st</sup> century. *Educational Research Review*, *13*, 74–83.

- Greiff, S., Scheiter, K., Scherer, R., Borgonovi, F., Britt, A., Graesser, A., Kitajima, M., & solving, Rouet F. Adaptive problem (2017). *Moving towards a new assessment domain in the second cycle of PIAAC, OECD Education Working Papers No. 156*. Paris: OECD Publishing.
- Guimond, S., Branscombe, N. R., Brunot, S., Buunk, A. P., Chatard, A., Désert, M., ... Yzerbyt, V. (2007). Culture, gender, and the self: Variations and impact of social comparison processes. *Journal of Personality and Social Psychology, 92*, 1118–1134.
- Guiso, L., Monter, F., Sapienza, P., & Zingales, L. (2008). Culture, gender and math. *Science, 320*, 1164–1165.
- Halpern, D. F. (2012). *Sex differences in cognitive abilities*. New York: Psychology Press.
- Halpern, D. F., Benbow, C. P., Geary, D., Gur, R., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1–51.
- Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229–246.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269*, 41–45.
- Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology, 28*, 1010–1026.
- Huguet, P., & Regner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*, 1024–1027.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.
- Iervolino, A. C., Hines, M., Golombok, S. E., Rust, J., & Plomin, R. (2005). Genetic and environmental influences on sex-typed behaviour during the preschool years. *Child Development, 76*, 826–840.
- Ikenaga, T., & Kambayashi, R. (2010). *Long-term trends in the polarization of the Japanese labor Market: The increase of non-routine task Input and its valuation in the labor market*. Hitotsubashi University Institute of Economic Research Working Paper.
- Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport: Praeger.
- Johnson, D. J., & Wilson, J. P. (2019). Racial bias in perceptions of size and strength: The impact of stereotypes and group differences. *Psychological Science, 30*, 553–562.
- Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. New York, NY: Routledge.
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. Oxford: Oxford University Press.
- Lakin, J. M. (2013). Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence, 41*, 263–274.
- Lawton, C. A., & Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex Roles, 53*(9–10), 717–725.
- Lindberg, S., Hyde, J., Petersen, J., & Linn, M. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*, 1331–1332.
- Mayer, R. E., & Wittrock, R. C. (2006). Problem solving. In P. A. Alexander, & P. H. Winne (Eds.). *Handbook of Educational Psychology*(2nd ed.). Mahwah, NJ: Erlbaum.
- Meisenberg, G. (2016). Gender differences in school achievement across cultures: An analysis of results from PISA 2000–2012. *Mankind Quarterly, 57*, 227–251.
- Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*, 37–45.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives, 24*, 129–144.
- Nolleberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review: Papers and Proceedings, 106*, 257–261.
- OECD (2014a). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems*. Paris: OECD Publishing.
- OECD (2014b). *PISA 2012 technical report*. Paris: OECD Publishing.
- OECD (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. Paris: OECD Publishing.
- OECD (2016). *PISA 2015 results. Excellence and equity in education. Volume I*. Paris: OECD Publishing.
- Perez-Felkner, L., McDonald, S., Schneider, B., & Grogan, E. (2012). Female and male adolescents' subjective orientations to mathematics and the influence of those orientations on postsecondary majors. *Developmental Psychology, 48*, 1658–1673. <https://doi.org/10.1037/a0027020>.
- Reilly, & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles, 68*, 521–535.
- Rindermann, H. (2007). The g-Factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*, 667–706.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*, 142–151.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.
- Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics, 24*, 235–270.
- Steele, C. M. (1997). A threat is in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.
- Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PLoS ONE, 11*.
- Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence, 48*, 137–151.
- United Nations Development Programme (UNDP) (2016). *Human development report 2016 human development for everyone*. New York: United Nations Development Programme.
- Valla, J., & Ceci, S. J. (2011). Can sex differences in science be tied to the long reach of prenatal hormones?: Brain organization theory, digit ratio (2D/4D), and sex differences in preferences and cognition. *Perspectives on Psychological Science, 6*, 134–146.
- VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. Posner (Ed.). *Foundations of cognitive science*. Mahwah, NJ: Erlbaum.
- Vuoksima, E., Kaprio, J., Kremen, W. S., Hokkanen, L., Viken, R. J., Tuulio-Henriksson, A., & Rose, R. J. (2010). Having a male co-twin masculinizes mental rotation performance in females. *Psychological Science, 21*, 1069–1071.
- Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*, 412–423.
- Weinberg, B. (2000). Computer use and the demand for female workers. *Industrial and Labor Relations Review, 53*, 290–308.
- Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science, 21*, 258–262.
- Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29*, 18–29.