Social Psychology of Education (2020) 23:1103–1119
https://doi.org/10.1007/s11218-020-09545-3

Does cognitive reflection mediate the math gender gap at university admission in Chile?

Jairo A. Navarrete[1] · **José S. Sandoval-Díaz**[2]

Received: 14 June 2019 / Accepted: 31 January 2020 / Published online: 27 March 2020

Abstract

In Chile, a vast and persistent gender gap in math performance at university admission has negative consequences for women's opportunities. International evidence suggests that these gender differences reflect gender inequities in educational and economic opportunities available in a given culture. A theoretical model suggests that sociocultural factors such as gender stereotype and math anxiety have an impact on women's math abilities by affecting their skills for cognitive reflection. This cross-sectional study collected data from 259 university students to gather preliminary evidence on the close relationship between gender gaps in math achievement and gender gaps in cognitive reflection. A mediation analysis shows that the level of cognitive reflection of students fully mediates the effect of gender on math achievement, even after controlling for participants' linguistic abilities. These findings support the possibility that there is a close relationship between gender differences in cognitive reflection and gender gaps in math achievement at university admission in Chile. The discussion emphasizes that the university admission's data is under the control of state agencies with strict protocols of information access that undermines the feasibility of researching this area. Assessments of cognitive reflection are easy to administer in only a few minutes and under laboratory conditions. Hence, one way to boost the viability of research and diagnosis of gender gaps in Chile is to use cognitive reflection assessments as proximal estimates for gender gaps in math performance at university admission.

Keywords Cognitive reflection · University admission · Math achievement · Gender gap

✉ Jairo A. Navarrete
jairo.navarrete@uoh.cl

José S. Sandoval-Díaz
jsandoval@ubiobio.cl

[1] Instituto de Ciencias de la Educación (ICEd), Universidad de O'Higgins, Av. Libertador Bernardo O'Higgins 611, Rancagua, Chile

[2] Centro de estudios Ñuble y Grupo de Investigación Género, Ciudadanía y Equidad (GI 170424/EF), Universidad del Bío-Bío, Chillán, Chile

1 Introduction

International reports signal gender differences regarding fundamental rights such as health, education, economic participation, and political empowerment across nations (Hausmann et al. 2008). A prominent example of gender inequity across countries appeared in the STEM curriculum, designed to foster social innovation with a focus on educating professionals in science, technology, engineering, and mathematics. Across the 27 countries of the European Union, females constitute less than 20% of students enrolled in the two dominant STEM careers of computing and engineering (European Commission et al. 2015). In the United States, only 18% of the undergraduate students enrolled in engineering programs are women (NSF 2013; Hill et al. 2010). In Chile, women account for only 21% of students in the entire field of technology (CNED 2012). To counter gender inequity, it is essential to improve this situation as education fosters personal development and facilitates the achievement of individuals' economic stability and security.

Chile has one of the most significant gender gaps in math achievement worldwide (OECD 2015), only below Colombia and Luxembourg. In Chile, university admission strongly depends on a battery of standardized tests (PSU by their Spanish acronym) in the areas of mathematics and language. Since the first PSU application in 2003, gender gaps in math performance have been increasing: from 2013 to 2016, the gender gap was about one-fourth of a standard deviation ($d = 0.25$). In 2017, it is remarkable that 127 male students attained the maximum score in mathematics, whereas only 24 female students achieved the same result. Unlike in other countries, where admission involves a comprehensive review of students' background, admission to Chilean universities rely strongly on PSU assessments (since 2013, admission also considers high-school rank). Since PSU tests bear life-lasting consequences for test-takers, the gender gap in math achievement implies that women have fewer opportunities to select career paths and to obtain state benefits such as scholarships. This status-quo undermines women's future income, social status, and socioemotional well-being (Bynner and Parsons 1997; OECD 2012; Parsons and Bynner 2005).

The problem, however, seems like the old chicken-egg causality dilemma. On the one hand, the above discussion outlines how gender gaps in math achievement are translated into gender socio-economic inequity as they promote underrepresentation of women in STEM careers but also in economic, political, and academic leadership (OECD 2012, 2013). On the other hand, a meta-analysis reveals that factors such as gender equity in school enrollment, women's share of research jobs, and women's parliamentary representation predict cross-national variability of gender gaps in math achievement (Else-Quest et al. 2010). This dilemma highlights the complexity of the phenomenon that entangles socio-cultural factors and gender gaps in math achievement. Therefore, it is necessary to refer to a conceptual model that facilitates the understanding of the situation. This work will propose a theoretical model of how socio-cultural factors nurture gender gaps in math achievement. The next section provides a theoretical framework and discusses some of the factors that are associated with gender differences in math achievement.

1.1 Sociocultural and psychosocial factors nurturing gender gaps in math achievement

The fact that both cognitive abilities and non-cognitive abilities influence math achievement (Cunha and Heckman 2007), makes it challenging to identify the genesis of gender gaps in math performance. Many educational systems worldwide display gender gaps, but these gaps are not systematic since countries such as Finland and Norway do not show gender gaps in math achievement (OECD 2015). Furthermore, according to a recent meta-analysis, a gender gap in math achievement observed in the US in 2009 has disappeared recently (Hyde 2016), thus pointing out that socio-cultural and psycho-social factors (rather than biological ones) nurture gender gaps. The purposes of this study make it necessary to provide a short description of some influential factors acknowledged in the literature of gender gaps in math achievement.

S*tereotype threat* refers to the experience of one being in a situation where one faces judgment (from oneself and others) based on societal stereotypes about one's group. When a group's stereotype judges a critical ability, individuals of such a group diminish their performance on activities which depend on such an ability, especially when there are judgments about this ability. Spencer et al. (1999) explained that such diminishing is due to performance being under the extra pressure of being judged by the stereotype. Our particular concern here is gender stereotype, which refers to the threat affecting women as a group (Anker 1997; Espinoza et al. 2014; OECD 2012, 2015; Spencer et al. 1999). There is comprehensive evidence about the negative effect on woman's math performance of being stereotyped as poor math performers (Ramsey and Sekaquaptewa 2011; Smith and Hung 2008). Empirical evidence shows that stereotype threat has a substantial impact on the likelihood of women leaving STEM careers (Beasley and Fischer 2012). Additionally, gender stereotype disrupts women's math performance at early ages: math performance of girls whose mothers endorse gender stereotype decline from kindergarten to 2nd grade (Tomasetto et al. 2011). This effect also applies to teachers providing social cues that support gender stereotypes (Beilock et al. 2010).

When a woman develops a negative perception of her abilities in mathematics, the negative effect of stereotype threat on women's math performance increases (Zimmerman 2000). In this regard, *self-efficacy* is a concept that comprises personal judgments of one's capabilities to organize and execute courses of action to attain designated goals, such as obtaining excellent math performance. It embodies the wise suggestion of "believe in yourself" given by experienced teachers to students. Self-efficacy is a strong predictor of students' motivation and learning (Bandura 1997; Pajares and Miller 1994; Zimmerman 2000), and empirical evidence shows that men have higher levels of math self-efficacy than women (Pajares and Miller 1994). For example, a national survey in the US, with a focus on mainstream college calculus, revealed that when comparing women and men with above-average mathematical ability, women start and end the term with significantly lower mathematical confidence than men. This evidence suggests that a lack of mathematical self-efficacy, rather than a lack of mathematical ability, is responsible for the high proportion of women that abandon pursuing STEM careers. The study points out that the odds of a woman desisting from continuing in calculus are 1.5 times greater than the odds for a man (Ellis et al. 2016).

Factors like the ones described above likely raise feelings of anxiety and helplessness. In particular, *math anxiety* increases in the presence of challenging mathematical situations and is characterized by an acute fear of mathematics that disrupts cognitive strategies and working memory (Ashcraft 2002). There exists a well-established negative correlation between math anxiety and math achievement, and this correlation is more pronounced in children with the highest potential (Ramirez et al. 2013). Math anxiety impairs healthy development of cognitive abilities such as numeracy (Ashcraft and Krause 2007), even in children as young as 7 years old (Ramirez et al. 2013). There is a consensus that math anxiety undermines numeracy skills and math performance by overloading on-line working memory, which is burdened by negative thoughts and negative ruminations about student's helplessness to achieve math success (Ashcraft and Krause 2007). International reports have highlighted that school-related anxiety is more prevalent in women (OECD 2016).

1.2 A model for the cognitive mechanisms underlying math gender gaps

This section aims to sketch how the sociocultural and psychosocial factors mentioned above can influence cognitive skills that determine math performance at university admission (MPUA). Above all, this section aims to underline the close relationship between math performance and a specific cognitive ability called *cognitive reflection*. More precisely, this study proposes that this cognitive ability embodies a bridge between psycho-socio-cultural factors and MPUA.

The *cognitive reflection test* (CRT) measures a prominent trait in decision-making that assesses the ability of a problem-solver to inhibit the first answer that comes to mind in the benefit of providing a more deliberative response (Frederick 2005). Moreover, researchers have argued that assessments of cognitive reflection measure rational thinking and analytic cognitive style (Shah et al. 2017; Toplak et al. 2011, 2014). It is necessary to emphasize strong evidence regarding the close relationship between mathematical abilities and CRT-performance (see Weller et al. 2013). Finucane and Gullion (2010) reported a moderate correlation between numeracy and CRT ($r = .53$). Obrecht et al. (2009) showed that CRT-performance correlates with SAT quantitative scores ($r = .45$). Liberali et al. (2012) showed a correlation between the Lipkus scale and the CRT ($r = .51$ and $r = .40$ in the Brazilian and US samples, respectively). This evidence emphasizes the strong links between cognitive reflection and math performance (see path F in Fig. 1). The remaining of this section outlines how certain socio-cultural factors might nurture gender gaps in math performance by affecting cognitive reflection abilities differently in men and women.

Central to this work is the observation that various studies around the world have reported gender differences in CRT-performance (Campitelli and Gerrans 2014;

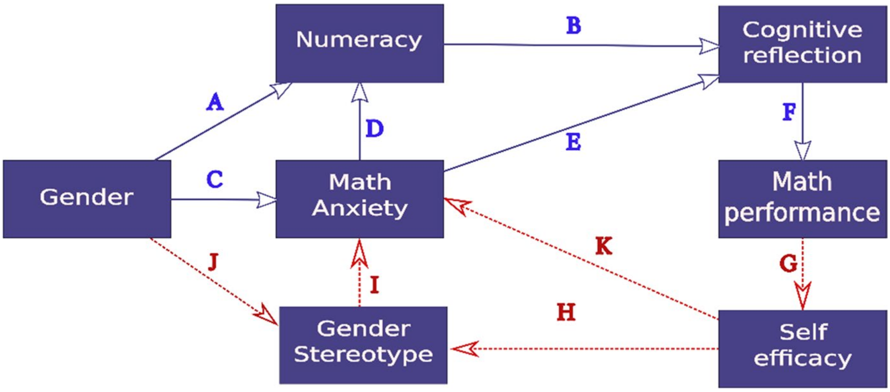

Fig. 1 Conceptual diagram that illustrates connections between psychosocial and sociocultural factors that nurture gender gaps in math performance at university admission processes

Frederick 2005; Primi et al. 2018). Prior research has shown that math reasoning abilities and anxiety feelings toward mathematics fully mediate gender gaps in cognitive reflection. Figure 1 depicts how numeracy abilities and math anxiety mediate the influence of gender on cognitive reflection abilities. Remarkably, the path following first math anxiety and then numeracy (CDB-path) was reported to be particularly robust (strength = .76; Primi et al. 2018), and the path through math anxiety (CE-path) was also significant (strength = .15). On the one hand, previous investigations support continuous lines in Fig. 1, thus suggesting that cognitive reflection mediates the effect of gender on math performance: gender exerts an influence on math performance indirectly through first affecting cognitive reflection abilities. On the other hand, the dotted lines in Fig. 1 illustrate how socio-cultural factors might affect MPUA indirectly: through the development of math anxiety or the undermining of numerical skills that ultimately weaken abilities for cognitive reflection.

There is no complete understanding of the interactions between the elements illustrated in Fig. 1, but some of the links are clear. Being targeted by gender stereotypes and feelings of low self-efficacy are triggers to math anxiety (Spencer et al. 1999; Bandura 1997). Math anxiety interferes with the healthy development of numeracy and cognitive reflection abilities (Ashcraft et al. 2007; Morsanyi et al. 2014). Above all, if one assumes that the gender influence on MPUA is due to gender-differentiated effects of accumulated impacts of various socio-cultural and psychosocial factors, the above discussion points to the hypothesis that CRT-performance should track the impact of gender on MPUA. Specifically, this work aims to gather preliminary evidence regarding one component of the model depicted in Fig. 1: This study seeks to test whether CRT-performance mediates the effect of gender on MPUA.

The next section describes a cross-sectional study conducted to gather evidence supporting the assumption that the level of cognitive reflection between men and women can explain gender gaps in math performance at university admission in Chile. The results section presents data analysis based on a mediation model confirming that cognitive reflection abilities fully mediate gender effects on math achievement. The discussion sets these results in the Chilean context by describing the Chilean educational system and pointing out implications for future research and applications to tracking up to the genesis of gender gaps in math achievement.

2 Method

2.1 Cognitive and performance measures

2.1.1 The cognitive reflection test

The first Cognitive Reflection Test was developed by Frederick (2005) and comprises three problems whose correct solution does not demand complex calculations but requires the suppression of an "impulsive" solution that readily comes to mind. As an example, the first problem states "A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?" To give the correct answer of 5 cents, one must suppress the intuitive but incorrect response of 10 cents. The CRT predicts rational thinking over and above cognitive ability, even after controlling for intelligence, cognitive skills, thinking dispositions, and executive functions (Frederick 2005; Toplak et al. 2011). Thus, people with high CRT scores are categorized as more likely to engage in rational, analytic thinking (Shah et al. 2017). A recent study (Primi et al. 2016) used item response theory to show that the original CRT items have high discriminative power and that the CRT is well suited to be used as a population screen for the cognitive reflection trait. The study also evaluated the precision of this test at different levels of the measured construct and showed that the CRT scale is highly reliable within the range of the trait from $-.4$ to .4 standard deviations around the mean ($r = .84$). In this study, we used a more recent version of the cognitive reflection test known as the "long version" of the cognitive reflection test (CTRL) and developed by Toplak et al. (2014). It consists of seven problems, including three original ones.

2.1.2 The PSU test of mathematics

The Department of Assessment, Measurement, and Register of the University of Chile (DEMRE for its acronym in Spanish) is in charge of administering this national test. The mathematical component of the PSU (PSU-math) is a battery of tests assessing mathematical knowledge in four subjects: numbers, algebra, geometry, and basic statistics. It comprises a measure of mathematical reasoning that assesses cognitive abilities, operation modes, and general methods of problem-solving. It also evaluates the comprehension and application of mathematical knowledge as well as its applications to analysis and resolution of mathematical problems (DEMRE 2016a). The test contains 75 multiple-choice items administered to students in sheets of paper under supervision. This test has high reliability, as measured by a Cronbach's alpha of 0.923 (DEMRE 2017:33). The standardized scores have a minimum and a maximum of 150 and 850, respectively, whereas their mean and standard deviation are respectively 500 and 110 points.

2.1.3 The PSU test of language

DEMRE is in charge of administering this national test. The language component of the PSU testing (PSU-lang) is a battery of tests assessing reading comprehension of literary texts as well as press writing and expository texts. It comprises an assessment of reading comprehension and reading processes. It also evaluates the application of comprehension strategies, identification, and reflective analysis of characteristic features of texts and capabilities for assessing its quality. Additionally, it indirectly evaluates writing processes by determining the variety of lexicon, connectives, writing strategies, and appreciation of writing in its creative and reflexive dimensions (DEMRE 2016b). The test contains 75 multiple-choice items administered to students in sheets of paper under supervision. This test has high reliability, as measured by a Cronbach's alpha of 0.91 (DEMRE 2017:33). The standardized scores have a minimum and a maximum of 150 and 850, respectively, whereas their mean and standard deviation are respectively 500 and 110 points.

2.1.4 Participants and study design

This study has a cross-sectional design that controls language abilities as predictors of math performance. Table 1 shows the composition of the 259 undergraduates recruited for this study (157 females, 158 first-year students or sophomore) in terms of the careers they were pursuing (Commercial Engineering, Computational Engineering, Math Teacher, Psychology and Nursing) in a state university located in the south of Chile. The ethical committee of the university reviewed and approved the design of the study and the informed consent presented to participants. Participants had an average age of 20.8 years (SD = 2.6), and their socio-economic status is reported here by the maximum level of education completed by their mothers (3.5% none, 26.3% school, 43.2% high school, 16.2% technical education, 10.8% university education). Participants attended in small groups (size 15–20) to the computation laboratory of the university where they signed informed consents, received instructions about the study, and completed an online survey. The Qualtrics research software and its online platform delivered the online form. To prevent participants from hiding sensitive information, feeling anxious, or under evaluation, they completed

Table 1 Composition of the sample of 259 participants

	Math related programs			Social science programs	
	Comm. Eng. (%)	Comp. Eng. (%)	Math teach. (%)	Psych. (%)	Nursing (%)
Male	5.0	12.0	3.9	12.4	6.2
Female	7.3	2.7	3.9	23.2	23.6

Table 2 Correlation between targeted variables

	PSU-lang	PSU-math	CRTL-time
PSU-math	.43***	–	–
CRTL-time	.11	.21***	–
CRTL-Score	.25***	.40***	.41***

Significance: ***$p < .001$

Table 3 Gender differences on target variables

Variable	Women (N = 157)		Men (N = 102)		Gender differences		
	M	SD	M	SD	95% CI	*P* value	Effect size (d)
PSU-math	562.64	64.52	590.75	62.29	(12.11, 44.1)	< 0.001	0.44
CRTL-time	343.4	123.1	403.28	147.38	(25.18, 94.58)	< 0.001	0.44
CRTL-Score	0.94	1.22	2.26	10.9	(0.91, 1.75)	< 0.001	0.83

the survey anonymously. The study first asked participants for demographic information, collected additional information, and then administered the extended version of the cognitive reflection test. The software registered completion times for the cognitive reflection test. After delivering the cognitive reflection test, a final question appeared, namely, "have you seen any of these questions before?" intended to control for individuals familiar to the cognitive reflection test.

3 Results

3.1 Preliminary analyses

Preliminary analyses found moderate and statistically significant correlations between the target variables (see Table 2). In particular, the correlation between PSU-math scores and CRTL-scores is in line with international university admission processes across countries, for example, the math component of the SAT used in the United States correlates to CRT scores with a magnitude of $\rho = 0.46$ (Frederick 2005). Independent samples t-tests revealed gender gaps across some of the targeted variables (see Table 3). The gender gap of 28.2 points regarding PSU-math scores is consistent with national reports, namely 27.4 points in 2013 (DEMRE 2013), 29.9 points in 2014 (DEMRE 2014), and 27 points in 2015 (DEMRE 2015). The significant gender gap of 1.33 points in CRTL scores confirmed the gender gaps reported internationally (Frederick 2005; Campitelli and Gerrans 2014; Primi et al. 2018). Although there was a gender gap in completion times for the cognitive reflection test, and these completion times mildly correlated with PSU math scores, complementary regression analyses showed that completion times were not predictive of PSU math scores. Hence, the mediation analysis did not consider completion times.

In contrast, PSU language scores were significantly predictive of PSU math scores and used as covariates in the mediation analysis presented in the next section.

3.2 Mediation analysis

Many psychological studies aim to establish the effect of an independent variable X on a dependent variable Y by relying on techniques of analysis of variance or multiple regression. However, these techniques do not help to answer questions about the *process* whereby such effect takes place. In a mediation analysis, the statistical model includes a variable M that transmits the effect of X on Y to represent the process whereby the influence of X takes place on Y. Consequently, in this model, there is a *direct effect* of X on Y (i.e., X→Y), along with an *indirect effect* of X on Y mediated by M (i.e., X→M→Y). This indirect effect represents the process whereby X affects Y through M. The data supports the mediation model when the indirect effect is statistically significant, implying that the mediating variable is key to explain the effect of X on Y. For those readers interested in details about mediation analysis, a deeper discussion can be found in Hayes (2013) and MacKinnon et al. (2007).

To assess the effect of gender on MPUA, directly and via cognitive reflection, we conducted a mediation analysis that considers PSU-math scores as the outcome variable, gender as the independent variable and cognitive reflection as the mediation variable. The mediation model also included PSU language scores as a covariate to control for participants' linguistic skills. All continuous variables were normalized prior to performing the mediation analysis. We tested the significance of links in the mediation model depicted in Fig. 2 by using bootstrapping procedures. We computed standardized indirect effects for each of 50,000 bootstrapped samples and estimated the 95% confidence by determining the indirect effects at the 2.5th and 97.5th percentiles. If the bootstrapped confidence interval does not include zero, the corresponding effect is significant at $p < .05$ (see Preacher and Hayes 2008). Linguistic ability of participants (PSU-lang scores) significantly predicted level of cognitive reflection ($\beta = .19$, $p < .001$, $CI_{95} = [.07, .30]$) as well as PSU-math scores

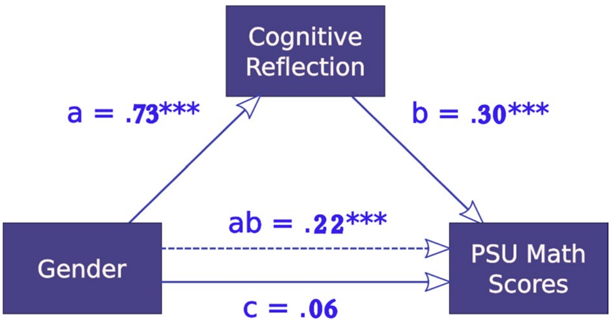

Fig. 2 Simple mediation model: cognitive reflection mediates the effect of gender on math achievement. The indirect effect of gender on MPUA through cognitive reflection is statistically significant (dotted arrow), whereas the direct influence of gender on MPUA is not significant (continuous arrow)

($\beta = .35$, $p < .001$, $CI_{95} = [.24, .46]$). After controlling for linguistic ability, the relationship between gender and cognitive reflection was significant ($a = .73$, $p < .0001$, $CI_{95} = [.50, .96]$) as it was the relationship between cognitive reflection and PSU math scores ($b = .30$, $p < .001$, $CI_{95} = [.18, .42]$). Above all, the bootstrapped standardized *indirect effect* was statistically significant ($ab = .22$, $p < .001$, $CI_{95} = [.13, .34]$), whereas the *direct effect* of gender on PSU-math scores was small and not statistically significant ($c = .06$, $p = .57$, $CI_{95} = [-.16, .30]$). Since gender did not affect PSU math scores when cognitive reflection mediates the relationship between them, cognitive reflection *completely mediates* the effect of gender on PSU math scores.

4 Discussion

The introduction of this work suggested that gender gaps in math achievement and gender gaps in cognitive reflection are closely related. This study gives a first step towards testing the conceptual model described in Sect. 1.2 (Fig. 1) by focusing on testing the central hypothesis, namely, that people's abilities for cognitive reflection are strongly associated with the math gender gap in Chile. Our data support that variations in students' abilities for cognitive reflection mediate gender gaps in math-performance at university admission. Prior research has shown that math reasoning abilities and anxiety feelings toward mathematics completely mediates the effect of gender on cognitive reflection. These elements outline a rationale where sociocultural and psychosocial factors exert indirect, gender-differentiated impacts on young people's abilities for cognitive reflection that eventually appear as gender gaps in math assessments at university admission. This explanation points out the possibility that factors originating the math gender gap at university admission in Chile have sociocultural and psychosocial roots such as gender stereotypes, women's sense of math self-efficacy, and women's propensity to math anxiety. Although the present research cannot explain the different social mechanisms whereby these key factors originate and reinforce gender gaps, the theoretical model (Sect. 1.2) paves the way for future research. For example, it seems essential to investigate the impacts of variables such as gender stereotype, math self-efficacy, and math anxiety on gender gaps (in either cognitive reflection or math performance). The following discussion aims to provide a more comprehensive interpretation of the presented results in the Chilean context and the international context.

4.1 Evidence from the Chilean educational system

In the Chilean context, it is unavoidable to consider the socioeconomic inequity of the educational system. The strength of the relationship between performance and socioeconomic status is one of the highest in the (OECD 2013), and thus high-quality education is associated with high-income families (a large portion of the educational system is under private control at the secondary and tertiary level). This educational inequality is at the roots of large socioeconomic inequality and segregation (Valenzuela et al. 2014). University diplomas are strongly pursued by young people as they have high socioeconomic value: Chile is the country with the largest difference in relative incomes between tertiary and non-tertiary graduates (OECD 2013). In this context, the socioeconomic status seems to be important to explain the gender gap in math achievement. However, to the extent of our knowledge, there is no evidence supporting such a hypothesis. Moreover, a prior study analyzed large longitudinal datasets from national agencies in charge of applying tests and surveys that measure the quality of education and found that socioeconomic factors do not help to explain the math gender gap in Chile (Bharadwaj et al. 2016). The study assessed the influence of socioeconomic factors such as parents' background characteristics, parents' investment behavior, students' classroom environment, and students' individual characteristics (e.g., self-assessed math ability) on the math gender gap. Despite the huge size of the dataset (more than a million observations), the researchers concluded that none of these factors accounted for a substantial portion of the math gender gap, even after controlling for school and class composition. In particular, sorting across schools or classes does not affect the math gender gap. The alternative view given by the present study provide evidence supporting a qualitatively different explanation of the math gender gap based on psychosocial and sociocultural variables. This further highlights the difficulties of understanding the nature of the math gender gap.

In this sense, the main contribution of this study is to highlight that a quick test (easily deliverable under laboratory conditions) can assess gender-variations of math performance at national assessments granting access to university education. Notice that the high-stakes national math assessments are under the control of government agencies with fixed annual schedules, strict protocols of access to information, low levels of flexibility, and high levels of bureaucracy. Therefore, it is worth contemplating the use of cognitive reflection tests as proxy measures to assess gender gaps in MPUA. This strategy might enhance the feasibility of research projects and interventional studies. For example, assume a research project aimed to investigate whether specific socioeconomic variables moderate gender gaps at MPUA. It might thus be wise first to conduct a low-effort pilot study to examine whether such socioeconomic variables moderate gender gaps in cognitive reflection. Performing this highly feasible pilot study would provide valuable information to lower the risks of an expensive research project whose logistics and schedule would depend mostly on external institutions.

4.2 Limitations and further research

This study collected participants' measures of cognitive reflection some 2 years after the national agency collected their PSU measures. Although this is a limitation, cognitive reflection is a stable trait that remains unchanged during long periods: a study on a large sample of participants ($N > 3000$) reported a strong correlation ($r = .755$) between two measures of cognitive reflection taken at two points in time separated by 2 years (Stagnaro et al. 2018). Still, it is possible that 2 years of university education had promoted cognitive reflection abilities, and then this issue should be addressed in future research through stricter methodologies controlling for the time-delay between CRT and PSU measures. In the meantime, we confirmed our results by repeating the mediation analysis but using only data belonging to students who enrolled in the university the year before conducting this study (61% of the sample, N = 156). This mediation analysis yielded comparable results to the ones reported before in terms of the statistical significance of the links in Fig. 2.

Regarding mediation analysis, specialized literature suggests that causal relationships of X on Y considering the variable M as a mediator of such influence should be inferred when X represents random assignments to conditions (i.e., an experimental design). This argument is based on that X precedes temporally to M and Y. In contrast, less restrictive designs (e.g., cross-sectional studies) need to consider the *equivalent model criticism:* alternative models might explain the same data equally well (for example, that X is the mediator of the relationship Y → M or that both M and Y cause X). This is due to that it is not always possible to distinguish between the alternative models without additional information (MacKinnon et al. 2007). In the context of the present study, researchers should not expect an experimental design since no design could randomly assign gender to participants. Nevertheless, some considerations improve the likelihood of a causal relationship for the present case. In the first place, gender is an independent variable that clearly precedes Cognitive Reflection and MPUA. In the second place, variables such as cognitive reflection and MPUA cannot have effects on gender (taken as a biological factor). Hence, there is only one alternative mediating model that competes with the one presented in this study: the one that considers MPUA as a mediator of the effect of Gender on Cognitive Reflection. Our data offer no support for such a mediation model: the direct effect of gender on cognitive reflection is large and strongly significant, whereas the indirect effect (mediated by MPUA) is small and statistically non-significant.

An aspect to cover in future research is the ongoing debate about the specific trait measured by tests of cognitive reflection. Classical theories posit that cognitive reflection measures the tendency of a problem-solver to inhibit the first intuitive answer that comes to mind (Frederick 2005), but recent views propose that cognitive reflection assesses rational thinking (Shah et al. 2017), or miserly processing of information (Toplak et al. 2014). Indeed, CRT-performance has been linked to the mathematical ability to such an extent that there are even claims that cognitive reflection is only a test of numerical competence (see Weller et al. 2013). This study does not share this latter view. We acknowledge that further research is necessary to increase our understanding of the cognitive reflection trait and its links to math performance.

Since the theoretical frame of this study draws on international research, it might be likely that cognitive reflection mediates MPUA in other countries with gender gaps in math achievement. However, international research is necessary to confirm these assertions. Future research might also contribute by further elaborating and testing the conceptual model described in Fig. 1. This study has given the first step to shed some light on the role of cognitive reflection, which acts as a hub that connects sociocultural and psychosocial factors with mathematics performance at university admission. However, the conceptual model might be useful to build novel hypotheses or relationships regarding gender gaps at MPUA. Notice, for example, that the diagram emphasizes the existence of self-reinforcing negative loops. Let us illustrate this point by giving an interpretation to the closed path GHIDBF: when a woman receives a poor grade, it becomes physical evidence that her performance is below the expected level. This grade likely affects her sense of math self-efficacy (G) since her self-beliefs in math abilities accommodate the evidence. When her parents, siblings, peers, and teachers acknowledge the poor grade, the gender stereotype on her environment becomes stronger (H), thus making math anxiety to rise (I). Carrying these feelings for a prolonged period might impair her development of numeracy (D) and cognitive reflection (B) abilities. These diminished cognitive reflection abilities influence her next grade (F), thus making this pattern reinforce itself and replicate countless times. The loop becomes stronger and more challenging to break out. This interpretation is consistent with prior research: math self-efficacy and math performance reinforce each other at developing gender differences (Sewasew et al. 2018).

4.3 Educational implications

Our results indicate that cognitive reflection acts as a hub that connects sociocultural and psychosocial factors with MPUA. We already mentioned that prior research has shown that the most robust connection between math performance and cognitive reflection is via feelings of anxiety. Figure 1 suggests that math-anxiety is central to the model in the sense that all the paths leading to MPUA must step through the math-anxiety node. Although the same considerations hold for cognitive reflection, in practice, it is better understood how to lower mathematical anxiety than how to promote cognitive reflection. Consequently, one strategy to overcome gender gaps in MPUA might be preventing the development of math anxiety in children and young people. Chile not only has one of the more significant gender gaps worldwide in math achievement (OECD 2015), but it also belongs to the ten OECD countries with the highest indexes regarding math anxiety in students, which is worsened by a strong negative association between high math anxiety and poor math performance (PISA 2013:100–102). The early bolstering of fundamental numerical and spatial processing skills in children has proven to be useful to reduce the likelihood of developing math anxiety (see Maloney and Beilock 2012). Some recently developed teaching strategies, whose design is based on aligning the learning materials' spatial structure with the numerical representations intended for children to learn (Navarrete and Dartnell 2017; Ramani and Siegler 2008), have been shown to promote numerical knowledge after children played with it for less than 2 h (Navarrete et al. 2018; Ramani et al. 2012). There is an increasing number of effective alternatives for teachers and schools to promote numerical knowledge in children through playful experiences, among them there is "The Great Race" (Ramani and Siegler 2008; Ramani et al. 2012) and software that improves young children's number sense with promising results in 7-to 9-year-olds suffering from dyscalculia (Wilson et al. 2009).

5 Conclusions

Investigating gender gaps in math performance at university admission is difficult due to the entanglement of socioeconomic, sociocultural and psychosocial factors that possibly explain its appearance. This kind of research is also challenging due to that admission processes are generally under the control of government agencies that impose multiple constraints on obtaining MPUA data. However, this study points out that measuring gender gaps on cognitive reflection is a promising indirect assessment of gender gaps on MPUA. This finding might foster research on gender effects on MPUA due to its applicability to lower the risk of performing expensive research projects. Further research is necessary to increase our understanding of the cognitive reflection trait and its relationship to MPUA. Since the theoretical underpinnings of this work are not specific to Chile, international research might help to understand whether the phenomenon reported here generalizes to other countries with gender gaps in math achievement.

Acknowledgements This research was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) del Gobierno de Chile, by funding of Program FONDEF Grant ID19I10030.

References

Anker, R. (1997). Theories of occupational segregation by sex: An overview. *International Labour Review, 136,* 315.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181–185.

Ashcraft, M. H., Jeremy, A. K., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability?. In *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 329–48). Baltimore: Paul H Brookes Publishing.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W H Freeman/Times Books/Henry Holt & Co.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education, 15*(4), 427–448.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860–1863.

Bharadwaj, P., de Giorgi, G., Hansen, D., & Neilson, C. A. (2016). The gender gap in mathematics: Evidence from Chile. *Economic Development and Cultural Change, 65*(1), 141–166.

Bynner, J. M., & Parsons, S. (1997). *Does numeracy matter?: Evidence from the National child development study on the impact of poor numeracy on adult life*. London: Basic Skills Agency.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*(3), 434–447.

CNED (Consejo Nacional de Educación) (2012). *Estadisticas de género Matricula primer año*, Chile.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47.

DEMRE. (2013). *Compendio estadístico proceso de admisión año académico 2013*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

DEMRE. (2014). *Compendio estadístico proceso de admisión año académico 2014*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

DEMRE. (2015). *Compendio estadístico proceso de admisión año académico 2015*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

DEMRE. (2016a). *Prueba de selección Universitaria. 1*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

DEMRE. (2016b). *Temario lenguaje de La prueba de Selección Universitaria*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

DEMRE. (2017). *Resultados batería PSU: Proceso admisión 2018*. Santiago de Chile: Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile.

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLoS ONE, 11*(7), e0157447.

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.

Espinoza, P., da Luz Fontes, A. B., & Arms-Chavez, C. J. (2014). Attributional gender bias: Teachers' ability and effort explanations for students' math performance. *Social Psychology of Education, 17*(1), 105–126.

European Commission, Directorate-General for Education and Culture, Dansk Teknologisk Institut, Technopolis Limited (Brighton), 3 s Unternehmensberatung GmbH, and ICF Consulting Services. (2015). *Does the EU need more STEM graduates? Final report*. Luxembourg: Publications Office.

Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging, 25*(2), 271–288.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Hausmann, R., Tyson, L. D. A., Zahidi, S., & World Economic Forum. (2008). *The global gender gap report 2008*. Geneva: World Economic Forum.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. ERIC.

Hyde, J. S. (2016). Sex and cognition: Gender and cognitive functions. *Current Opinion in Neurobiology, 38,* 53–56.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment: Numeracy in probability judgment. *Journal of Behavioral Decision Making, 25*(4), 361–381.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*(1), 593–614.

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences, 16*(8), 404–406.

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions, 10*(1), 31.

Navarrete, J., & Dartnell, P. (2017). Towards a category theory approach to analogy: Analyzing re-representation and acquisition of numerical knowledge. *PLoS Computational Biology, 13*(8), e1005683.

Navarrete, J. A., Gómez, D. M., & Dartnell, P. (2018). Promoting preschoolers' numerical knowledge through spatial analogies: Numbers' spatial alignment influences its learning. *Contemporary Educational Psychology, 54,* 112–124.

NSF. (2013). Women, minorities, and persons with disabilities in science and engineering: 2013: (558442013-001).

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition, 37*(5), 632–643.

OECD. (2012). *Closing the gender gap*.

OECD. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD.

OECD. (2015). *The ABC of gender equality in education*. Paris: OECD Publishing.
OECD. (2016). *PISA 2015 Results* (Vol. III). Paris: OECD.
Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193.
Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?*. London: National Research and Development Centre for Adult Literacy and Numeracy.
PISA. (2013). *PISA 2012 results* (Vol. III). Paris: OECD.
Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In *The SAGE sourcebook of advanced data analysis methods for communication research* (pp. 13–54). 2455 Teller Road, Thousand Oaks, California, 91320, United States: Sage Publications, Inc.
Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking & Reasoning, 24*(2), 258–279.
Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT): A new version of the cognitive reflection test applying IRT. *Journal of Behavioral Decision Making, 29*(5), 453–469.
Ramani, G., & Siegler, R. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*(2), 375–394.
Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology, 104*(3), 661–672.
Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development, 14*(2), 187–202.
Ramsey, L. R., & Sekaquaptewa, D. (2011). Changing stereotypes, changing grades: A longitudinal study of stereotyping during a college math course. *Social Psychology of Education, 14*(3), 377–387.
Sewasew, D., Schroeders, U., Schiefer, I. M., Weirich, S., & Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology, 54,* 55–65.
Shah, P., Michal, A., Ibrahim, A., Rhodes, R., & Rodriguez, F. (2017). What makes everyday scientific reasoning so challenging?. In *Psychology of learning and motivation* (vol. 66, pp. 251–99). Elsevier.
Smith, C. S., & Hung, L. C. (2008). Stereotype threat: Effects on education. *Social psychology of education, 11*(3), 243–257.
Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28.
Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive Reflection test is stable across time. *Judgment and Decision Making*, *13*(3), 260–267.
Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47*(4), 943–949.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275–1289.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning, 20*(2), 147–168.
Valenzuela, J. P., Bellei, C., & Ríos, D. D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy, 29*(2), 217–241.
Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A rasch analysis approach: Rasch-based numeracy scale. *Journal of Behavioral Decision Making, 26*(2), 198–212.
Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education, 3*(4), 224–234.
Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Jairo A. Navarrete is Academic-Researcher at the Institute of Educational Sciences at the Universidad de O'Higgins and serves as the Head of the Laboratory of Learning and Games at the same university. His research interests are in areas of cognition aimed to understand the role played by analogy making, relational thinking, and emotion in the learning of mathematics.

José S. Sandoval-Díaz Academic-Researcher at the Ñuble Studies Center and the Citizenship and Equity Research Group of the Department of Social Sciences, Universidad del Bío Bío, Chillan, Chile. Research interests: Educational vulnerability and socio-environmental risks.