Learning and Individual Differences 102 (2023) 102255

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Gender differences in young adults' mathematical performance: Examining the contribution of working memory, math anxiety and gender-related stereotypes

Helene Vos [a,*], Mila Marinova [b,c,d], Sara C. De Léon [e], Delphine Sasanguie [c,f], Bert Reynvoet [b,c]

[a] Faculty of Social and Behavioural Sciences, Institute of Education and Child Studies, Leiden University, Leiden, The Netherlands
[b] Faculty of Psychology and Educational Sciences, Research Unit Brain & Cognition, KU Leuven, Leuven, Belgium
[c] Faculty of Psychology and Educational Sciences @ Kulak, KU Leuven, Kortrijk, Belgium
[d] Faculty of Humanities, Education and Social Sciences, Institute of Cognitive Science and Assessment, Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-Belval, Luxembourg
[e] Faculty of Psychology and Speech Therapy, Department of Developmental and Educational Psychology, Universidad de La Laguna, La Laguna, Spain
[f] Research Centre for Learning in Diversity, HOGENT University of Applied Sciences and Arts, Ghent, Belgium

ARTICLE INFO

Keywords:
Gender differences
Mathematical performance
Working memory
Math anxiety
Gender-related stereotypes

ABSTRACT

Gender differences have been widely reported for mathematical performance tests such as basic arithmetic tests and more complex tests such as the cognitive reflection test. The current study examined which factors could explain these gender differences. Young adults (N = 189; 18–35 years) performed an arithmetic test and cognitive reflection test. Subsequently, it was examined to which extent gender differences on these tests could be explained by verbal and visuo-spatial working memory, explicit and implicit gender-related stereotypes and math anxiety. Results showed that women scored significantly lower than men on the arithmetic and cognitive reflection tests. A mediation analysis demonstrated that the relation between gender and arithmetic performance was partially mediated by math anxiety and explicit gender-related stereotypes. Furthermore, results showed that math anxiety fully mediated the relation between gender and cognitive reflection. These results demonstrate that math anxiety plays a key role in the relation between gender and mathematical performance.

1. Introduction

Nowadays, women are still underrepresented in science, technology, engineering and mathematics (STEM) education and careers. Worldwide, women comprise 35 % of students enrolled in higher-education STEM degrees (UNESCO, 2017). The underrepresentation of women in the STEM field has raised the discussion of whether the gender imbalance in STEM education is due to differences in mathematical abilities (Hyde et al., 2008; Wang & Degol, 2016). In general, previous findings have shown that gender differences in mathematical performance are not present in childhood (Bakker et al., 2019; Hutchison et al., 2019; Lachance & Mazzocco, 2006), but that differences favoring men emerge in high school and persist in adulthood (see Else-Quest et al., 2010; Lindberg et al., 2010 for meta-analyses). As mathematical ability comprises a wide variety of skills, gender differences in mathematical performance might be dependent on the test that is used (Hyde et al., 1990; Lindberg et al., 2010). For example, mathematical ability comprises arithmetic fluency, which refers to the usage of operations in an appropriate, efficient and flexible manner. In a meta-analysis, Lynn and Irwing (2008) examined gender differences in children and adults for the arithmetic subtest of the Wechsler Intelligence Scale. Results revealed that men performed better on mental arithmetic and that these gender differences emerged in late childhood and early adolescence. Since arithmetic fluency is the foundation for developing higher-order mathematical skills (Price et al., 2013), it might be no surprise that gender differences have also been found for more complex mathematical tests. Among one of these more complex mathematical tests are word problems which can be defined as verbal problem descriptions raising questions that can be answered by applying operations (Verschaffel et al., 2020). Whereas word problems are considered as complex mathematical tests, these problems are often solved by a shallow approach as one might select a suitable way to solve the problem on the basis of textual cues (Hickendorff, 2021). However, this shallow approach might not be applied to word problems that require reflection

* Corresponding author at: Institute of Education and Child Studies, Wassenaarseweg 52, 2333 AK Leiden, the Netherlands.
E-mail address: h.m.w.vos@fsw.leidenuniv.nl (H. Vos).

https://doi.org/10.1016/j.lindif.2022.102255
Received 11 August 2021; Received in revised form 18 December 2022; Accepted 23 December 2022
Available online 17 January 2023

on the problem structure and strategy. An example of such a test that requires reflection is the cognitive reflection test (CRT; Frederick, 2005). The CRT presents participants with mathematical word problems that bring an intuitive but incorrect answer to mind, while the correct answer requires further reflection. For example, one of the questions of the CRT is: 'A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?'. An intuitive answer to this question is 10 cents, while the correct answer is 5 cents. On the one hand, the CRT taps into a System 1, which is executed automatically and is driven by emotions. On the other hand, the CRT taps into a System 2 which requires conscious thought and effort. This is probably one of the reasons that the performance on the CRT is related to decision-making in real life situations (Juanchich et al., 2016: Toplak et al., 2017), since real-life situations also require to inhibit intuitive responses and further reflection to come to a correct answer. Studies have consistently reported that men perform better than women on the CRT and that this in turn affects differences in real-life outcomes (Cueva et al., 2016; Frederick, 2005; Primi et al., 2018). In sum, most mathematical tasks primarily assess the ability to understand and solve the problem, but the CRT also assesses the ability to suppress emotions. Therefore, the CRT is a highly relevant task to take into account when examining gender and negative attitudes towards mathematics.

Previous research has shown that several factors might play a role in the relation between gender and mathematical performance such as working, math anxiety and stereotypes (see Barroso et al., 2021; Peng et al., 2016, for meta-analyses). However, most studies often focus on one specific predictor or only include one specific mathematical test. Consequently, it remains unclear to which extent working memory, math anxiety and gender-related stereotypes combined, are related to mathematical performance. Furthermore, it remains also uncertain whether the contribution of working memory, math anxiety and gender-related stereotypes is different for basic and complex mathematical tests. Here, we take into account a basic task measuring arithmetic fluency and a complex mathematical task requiring the ability to suppress intuitive responses and further reflection on the problem. In the remainder of this introduction, we will outline previous research on working memory, math anxiety and gender-related stereotypes in relation to mathematical performance. In order to address the gaps in the literature, the current study aims to specify to which extent gender differences in arithmetic fluency and cognitive reflection are mediated by working memory, math anxiety and gender-related stereotypes.

1.1. Working memory

Working memory is commonly related to mathematical performance (DeStefano & Lefevre, 2004; Peng et al., 2016; Raghubar et al., 2010). Working memory is a system where verbal and spatial information can be stored and manipulated for a short period of time (Baddeley & Hitch, 1974; Engle, 2002; Miyake et al., 2000). The most well-known model of working memory is the multi-component model of Baddeley and Hitch (1974). This model assumes that working memory consists of a supervisory core system called the *central executive* (CE), which coordinates activities and information within the slave systems, namely the phonological loop and visuo-spatial sketchpad (Baddeley, 1996). The *phonological loop* (PL) is responsible for the temporary storage and rehearsal of auditory information (Baddeley et al., 1998), while the *visuo-spatial sketchpad* (VSSP) is responsible for the storage and rehearsal of visual and spatial information (Barton et al., 1995). Working memory is often studied by span tests requiring to keep in mind a series of items in a certain order. Forward recall tests, which require rehearsing information in the same order, measure the capacity of the PL or VSSP. Backward recall tests, which require recalling information in the reversed direction, measure the ability to recall and manipulate information. They are a measure of the interaction between the CE and the PL or VSSP.

Since working memory plays an important role in mathematical performance, this raises the question whether individual differences in working memory can explain gender differences in mathematical performance. Previous studies have investigated gender differences in working memory. For example, Lynn and Irwing (2008) found that the male advantage in mental arithmetic could not be explained by better performance on the Digit Span Test — a verbal working memory test where participants have to recall a sequence of words in the same or the reversed order. Similarly, Piccardi et al. (2019) found no gender differences with regard to the performance on a Digit Span Test, but it was found that males outperform females on a Corsi Block Test. In this visuo-spatial working memory test, participants see blocks being tapped and they subsequently have to tap the blocks in the same order. In line with these findings, several other studies have reported that men outperform women on visuo-spatial working memory tests, but not on verbal working memory tests (Kaufman, 2007; Pauls et al., 2013; Robert & Savoie, 2006). By contrast, Zilles et al. (2016) found that men performed better in both working memory domains. In conclusion, working memory plays a prominent role in mathematical performance. However, the results of previous studies examining gender differences in working memory have yielded mixed results and it is not clear to which extent working memory affects the relation between gender and mathematical performance.

1.2. Math anxiety

Numerous studies have consistently shown that math anxiety is strongly related to mathematical performance (Barroso et al., 2021; Beilock & Maloney, 2015; Caviola et al., 2022). Math anxiety refers to the fear, tension, and apprehension individuals experience when facing mathematics (Ramirez et al., 2018). Math anxiety is widespread and affects both children and adults. For instance, results of the Program for International Student Assessment (Organization for Economic Cooperation and Development, 2016) indicated that 33 % of the students experience helplessness while doing mathematics. Interestingly, about 14 % of the variation in mathematical performance can be explained by variations in math anxiety. It is often assumed that individuals with high levels of math anxiety are attending to both performing mathematics and their worries and, in turn, this negatively affects their mathematical performance (Ashcraft & Kirk, 2001; Eysenck et al., 2007). Consequently, individuals with high levels of math anxiety show worse performance on math achievement tests than the performance predicted solely by their mathematical abilities.

Several studies examining gender differences in math anxiety have found that women report higher levels of math anxiety than men (Hart & Ganley, 2020; Luttenberger et al., 2018; Suárez-Pellicioni et al., 2016; see Else-Quest et al., 2010 for a meta-analysis). This gender difference seems to increase with age (Hill et al., 2016). Likely, gender differences in math anxiety lead eventually to avoidance of math-related activities. For example, Jansen et al. (2016), conducted a Dutch nationwide study among adults and found that only for women, math anxiety was a significant mediator in the relation between math skills and the use of math in everyday life. These results suggest that math anxiety and the use of math in everyday life mutually influence each other, and gender plays a role in this relationship.

1.3. Gender-related Stereotypes

The view that men are better in mathematics is widely held and these stereotypes could affect females' mathematical performance (Ertl et al., 2017; Luttenberger et al., 2018). Women might experience stereotype threat resulting in impaired mathematical performance. For example, Beilock et al. (2007) tested two groups of women: a stereotype threat group, and a control group. In the former case, the participants were told that the study examined why women perform worse in mathematics. In the latter case, the participants were told that the study examined problem-solving. Results showed that the stereotype threat group performed worse than the control group, indicating the negative effects of stereotype threat on the women's cognitive performance. However, a meta-analysis of Flore and Wicherts (2015) showed that stereotype threat has only a small effect on females' math, science, and spatial skills.

Gender-related stereotypes can be measured both explicitly and implicitly (Gawronski & De Houwer, 2014). Examples of explicit measures are self-report questionnaires, where participants indicate whether they consider particular abilities as more masculine or feminine (Liben et al., 2002). Answers on explicit measures are the result of conscious thoughts about stereotypes. In contrast, implicit measures assess the unconscious thoughts about stereotypes (Greenwald & Banaji, 1995). An example of such an indirect measure is the Implicit Association Test (IAT, Greenwald et al., 1998), a test in which participants have to categorize words from different concepts (e.g., masculine and feminine words and words related to academic disciplines). Here, a distinction is made between stereotype-congruent (e.g., brother, mathematics) and stereotype-incongruent conditions (e.g., sister, mathematics). It is assumed that participants with stronger implicit gender-related stereotypes will respond more accurately and faster to a stereotype-congruent condition compared to a stereotype-incongruent condition. In a large-scale study, nation-level stereotypes about mathematics measured by the IAT were found to be strongly related to gender gaps in science and math achievement within countries (Nosek et al., 2009). Nosek and colleagues (2009) showed a relationship of implicit and explicit stereotypes with mathematics performance, but these relationships were opposing for men and women. Stronger associations between mathematics and males corresponded with more positive attitudes towards mathematics for men but with more negative attitudes towards mathematics for women. Since the stereotype that math is for males has been confirmed with both explicit and implicit stereotype measures, both implicit and explicit measures are important when examining gender-related stereotypes (Cvencek et al., 2011; Steffens & Jelenec, 2011; Vuletich et al., 2020).

1.4. The current study

As mentioned above, gender differences in mathematical performance have been widely reported. On the one hand, gender differences have been observed for basic computational tests such as arithmetic fluency (Lynn & Irwing, 2008). On the other hand, gender differences have been observed for higher order mathematical tests which measure reflection on the problem structure and strategy such as the cognitive reflection test (Cueva et al., 2016; Frederick, 2005; Primi et al., 2018). Furthermore, previous research has shown gender differences in working memory, math anxiety and gender-related stereotypes indicating that these factors might play a role in the relation between gender and mathematical performance. Previous research about the relation between working memory, math anxiety, gender-related stereotypes and mathematical performance has yielded mixed results (Caviola et al., 2022; Finell et al., 2022; Namkung et al., 2019). While a meta-analysis of Finell et al. (2022) showed a mediating role of working memory, other meta-analyses showed that the indirect effect of math anxiety on mathematical performance mediated by working memory was negligible (Caviola et al., 2022; Namkung et al., 2019). Nevertheless, research has consistently shown that women show more negative feelings towards math and that this in turn affects mathematical performance (Barroso et al., 2021; Else-Quest et al., 2010; Nosek et al., 2009). The relation between math anxiety, gender-related stereotypes and mathematics might be dependent on the type of math task. A meta-analysis has shown a stronger relation for advanced math domains compared to more foundational mathematics domains (Namkung et al., 2019).

In the current study, we examined which factors could explain gender differences in mathematical performance. First, we examined whether working memory, math anxiety and gender-related stereotypes explain unique variance in mathematical performance as measured by an arithmetic fluency test and cognitive reflection test. Second, we investigated whether the predictors that did explain unique variance could explain the relation between gender and mathematical performance. Based on previous findings, it was hypothesized that math anxiety and gender-related stereotypes mediate the relation between gender and mathematical performance. While this is probably the case for both arithmetic and cognitive reflection (Frederick, 2005; Lynn & Irwing, 2008), it might be especially the case for more complex problems such as cognitive reflection and to lesser extent for automatized problems such as arithmetic fluency.

2. Method

2.1. Participants

In total, 189 participants of Dutch nationality (90 men, 99 women, $M_{age} = 22.86$; $SD = 3.01$) participated in this study. The distribution of the participant's highest educational level completed was: 1.6 % VMBO (preparatory secondary vocational education), 15.7 % HAVO (senior general secondary education), 28.3 % VWO (university preparatory education), 9.4 % MBO (senior secondary vocational education), 19.4 % (higher professional education), 12 % University Bachelor's degree, 12.6 % University Master's degree. An independent samples *t*-test indicated that men and women did not differ significantly in terms of age, $t(187) = 1.45$, $p = .149$. A chi-square test showed that men and women did not differ significantly in terms of educational level, $\chi(6) = 1.73$, $p = .943$. Because there were no differences between men and women, age and educational level were not included in the analyses reported below. When age and educational level were considered as covariates in the analyses, results showed that age and educational level did not relate to the outcome variables. Furthermore, when age and educational level were included in the mediation analyses, the results did not change. The results of these mediation analyses can be found on the Open Science Framework: https://osf.io/8vwe7

As pre-registered, participants were excluded from the analyses when their accuracy score for a test deviated more than three standard deviations from the group mean. This resulted in the removal of the scores of 16 participants on one of the tests (we excluded seven participants for the Number Series Completion test, three participants for the Digit Span Forward test, two participants for the Digit Span Backward test, four participants for the Matrices Span Forward test, and three participants for the Implicit Association Test measuring implicit gender-related stereotypes). Table 1 displays the number of participants taken into account in the analyses for each test.

2.2. Procedure

The current study was pre-registered on AsPredicted. The pre-registered protocol is available at https://aspredicted.org/WMS_INO. The study was approved by the local ethics committee. Participants received a link to participate online. Beforehand, participants received information about the procedure followed by the instruction to read and sign the informed consent. Participants were asked to fill in demographic information about their gender, age, nationality and educational level. Subsequently, the presentation of the tests took place in the following order: Tempo Test Arithmetic, Digit Span Forward Test, Cognitive Reflection Test, Digit Span Backward Test, Matrices Span Forward Test, Number Series Completion Test, Matrices Span Backward Test, Abbreviated Math Anxiety Scale, Activity Questionnaire of the Occupations, Activity and Traits Questionnaire and Implicit Association Test.

2.3. Measures

2.3.1. Arithmetic test

An online version of the Tempo Test Arithmetic was administered

Table 1
Mean scores and the corresponding standard deviations for the measures.

					Men			Women			Total		
	t	*df*	*p*	Cohen's *d*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Arithmetic	4.51	187	<0.001	0.66	90	135.91	25.27	99	119.44	24.84	189	127.29	26.31
Cognitive reflection	2.44	187	0.015	0.35	90	2.10	1.02	99	1.72	1.13	189	1.90	1.09
Number series completion	2.31	184	0.022	0.34	88	12.19	2.20	94	11.40	2.41	182	11.79	2.33
PL	−0.12	184	0.907	0.02	89	9.93	2.20	97	9.97	2.03	186	9.95	2.11
PL * CE	−0.96	185	0.337	0.14	89	10.99	2.53	98	11.33	2.26	187	11.17	2.39
VSSP	1.65	183	0.101	0.24	89	7.01	2.38	96	6.49	1.91	185	6.74	2.16
VSSP * CE	1.00	187	0.316	0.25	90	7.92	3.49	99	7.42	3.33	189	7.66	3.40
Math anxiety	−4.70	176.21	<0.001	.68[a]	90	7.66	5.39	99	12.15	7.66	189	10.01	7.03
Explicit stereotypes (masculine)	2.52	187	0.013	0.36	90	5.74	4.96	99	4.09	4.05	189	4.88	4.57
Explicit stereotypes (feminine)	1.69	185	0.093	0.25	90	5.13	4.01	99	4.16	3.88	189	4.62	3.96
Implicit gender stereotypes	0.11	184	0.909	0.03	89	2.94	5.23	99	2.38	5.04	188	2.65	5.12

Note. PL = phonological loop, VSSP = visuo-spatial sketchpad, CE = central executive.
Males were coded as 0, Females as 1.
[a] Men reported less math anxiety than women, $t(176.21) = -4.70$, $p < .001$, $d = 0.68$ and this was the case for both Learning Math Anxiety ($t(168.85) = 3.89$, $p < .001$, $d = 0.56$ with $M = 2.07$, $SD = 2.67$ for men and $M = 4.01$, $SD = 4.11$ for women) and Math Evaluation Anxiety ($t(186.10) = -4.71$, $p < .001$, $d = 0.68$ with $M = 5.59$, $SD = 3.41$ for men and $M = 8.14$, $SD = 4.03$ for women).

(TTR; De Vos, 1992). The test consists of five columns: addition, subtraction, multiplication, division, and one with mixed operations consisting of forty items presented in increasing difficulty (De Vos, 1992). Participants had one minute to solve as many problems as possible in a column. Before the start of the test, instructions were displayed. When participants clicked a button to start the test, the timer started, and the addition column appeared on the screen for one minute. Participants had to type the answers with their keyboard. After one minute, a screen appeared informing participants that one minute had passed, and participants had to click a button to proceed with the subtraction column. This continued until the participants completed the column with mixed operations, after which the test was terminated. The raw score of the number of correct responses (ranging from 0 to 200) was used as an index of arithmetic performance. Cronbach's alpha reliability of the test was 0.98.

2.3.2. *Cognitive Reflection Test*

The Cognitive Reflection Test (CRT) measures the ability to override intuitively appealing but incorrect answers to arithmetic problems (Frederick, 2005). Three open-ended questions were displayed on the screen. Participants were informed that several problems were displayed varying in difficulty, and they received the instruction to answer as many problems as possible. The raw score of the number of correct responses (ranging from 0 to 3) was used as an index of cognitive reflection. Cronbach's alpha reliability for this translation of the CRT was 0.62.

2.3.3. *Number Series Completion Test*

The Number Series Completion Test measures the ability to detect the relations among items in a sequence. In the Number Series Completion Test, 14 sequences were presented consisting of several digits (minimum 4, maximum 6) and one question mark. The 14 sequences were displayed on the screen. Participants were informed that several sequences were displayed varying in difficulty and that their task was to type the number that would fit at the place of the question mark. The raw score of the number of correct responses (ranging from 0 to 14) was used as an index of Number Series Completion. Cronbach's alpha reliability for the test was 0.95.

2.3.4. *Digit Span Forward Test*

The Digit Span Forward Test (Wechsler, 2012) measures the phonological loop capacity. Before starting the test, instructions were displayed informing participants that a series of digits would be presented sequentially and that their task was to recall the sequence by typing the sequence with their keyboard. The test started with a practice trial, after which feedback was provided by giving the correct answer. Subsequently, the experimental blocks started. No feedback was provided during the experimental blocks. The test consisted of 8 blocks of 2 trials at each list length. The number of digits in the initial block was two and increased by one in each successive block up to a maximum of nine. Each digit was displayed for 1 s. The number of correctly recalled sequences (ranging from 0 to 16) was used as an index of working memory's phonological loop component (PL). Cronbach's alpha reliability of the test was 0.70.

2.3.5. *Digit Span Backward Test*

The Digit Span Backward Test (Wechsler, 2012) measures the interaction between the phonological loop and the central executive. Instructions informed participants that a series of digits would be presented sequentially. Their task was to recall the sequence in reverse order by typing the sequence with their keyboard. The test started with a practice trial, after which the correct answer was given. Subsequently, the experimental blocks started. No feedback was provided during the experimental blocks. The number of digits in the initial block was two and was increased by one in each successive block up to a maximum of eight. The number of correctly recalled sequences (ranging from 0 to 16) was used as an index of the interaction between the phonological loop component and the central executive (PL*CE). Cronbach's alpha reliability for the test was 0.73.

2.3.6. *Matrices Span Forward Test*

The Matrices Span Forward Test measures the visuo-spatial sketchpad capacity. The matrices presented were 3 × 3 grids with nine squares, of which one square was colored black. Before the start of the test, instructions were displayed informing participants that a series of matrices would be presented sequentially. Their task was to recall the sequence by indicating in which order the matrices were presented. The test started with a practice trial after which feedback was provided by giving the correct answer. Subsequently, the experimental blocks started. No feedback was provided during the experimental blocks. The test consisted of eight blocks of two trials at each list length. The number of matrices in the initial block was two and increased by one in each successive block up to a maximum of nine. The number of correctly recalled sequences (ranging from 0 to 16) was used as an index of the visuo-spatial sketchpad component (VSSP) of working memory. Cronbach's alpha reliability for the test was 0.67.

2.3.7. *Matrices Span Backward Test*

The Matrices Span Backward Test measures the interaction between the visuo-spatial sketchpad and the central executive. Again, 3 × 3 grids were presented with nine squares in which one square was colored black. Instructions informed participants that a series of matrices would be presented sequentially and that their task was to recall the sequence in reverse order. The test started with a practice trial after which the correct answer was given. Subsequently, the experimental blocks started. No feedback was provided during the experimental blocks. The number of matrices in the initial block was two and increased by one in each successive block up to a maximum of eight. The number of correctly recalled sequences (ranging from 0 to 16) was used as an index of the interaction between the visuo-spatial sketchpad and the central executive (VSSP*CE). Cronbach's alpha reliability for the test was 0.84.

2.3.8. Abbreviated Math Anxiety Scale

A translation of the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003) was administered. The AMAS (Hopko et al., 2003) measures math anxiety and contains nine items describing situations of encountering mathematics. Participants were asked to indicate how anxious they would feel in each situation on a 5-point Likert scale ranging from low anxiety to high anxiety. The questionnaire differentiates Learning Math Anxiety (5 items) and Math Evaluation Anxiety (4 items). The sum score of all the items was used as an index of math anxiety, with the total score representing the sum of the nine items (ranging from 0 to 36). Cronbach's alpha reliability for this translation of the AMAS was 0.90, with 0.80 for math learning anxiety and 0.87 for math evaluation anxiety.

2.3.9. Activity questionnaire

Explicit gender-related stereotypes were measured by the Activity Scale of the OAT-M (Liben et al., 2002). The scale consists of 25 items describing various daily activities (e.g., go to the beach). Participants were asked to indicate for each activity whether they considered the activity as an activity for men or women on a 5-point Likert ranging from "only men" to "only women". Liben et al. (2002) showed that 10 of these items are typically considered masculine (e.g., fly a model plane), 10 items are typically considered feminine (e.g., knit a sweater), and five items are considered neutral (e.g., go to the beach). The sum score of responses to the masculine items was used as an index of explicit gender-related stereotypes about men. The total score represented the sum of the ten items. The sum score of responses to the feminine items was used as an index of explicit gender-related stereotypes about women with the total score representing the sum of the ten items. Cronbach's alpha reliability was 0.92 for the masculine items and 0.87 for the feminine items.

2.3.10. Implicit Association Test

Implicit gender-related stereotypes were measured with an Implicit Association Test. The implicit association test presented here was a categorization test consisting of eight blocks in which participants have 10 s to categorize eight words in two frames. In line with Nosek et al. (2009), both male/female words (e.g., brother/sister) and science/liberal arts words (mathematics/history) were presented. In four congruent blocks, male and science words needed to be categorized in a frame, while female and liberal arts words needed to be categorized in another frame. In four incongruent blocks, female and science words needed to be categorized in a frame while male and liberal arts words needed to be categorized in another frame. The difference score between the accuracy scores of the congruent and incongruent blocks was used as an index of implicit gender-related stereotypes. Cronbach's alpha reliability for the test was 0.89.

2.4. Data-analysis

Table 1 presents the mean scores for the measures and *t*-tests examining whether gender differences were present. To examine whether working memory, math anxiety and gender-related stereotypes could account for gender differences in mathematical performance, several analyses were conducted. First, we conducted exploratory regression analyses to investigate whether working memory, math anxiety and gender-related stereotypes explain unique variance in mathematical performance as measured by the Tempo Test Arithmetic and Cognitive Reflection Test. Subsequently, the pre-registered main analyses were conducted. To examine which factors could account for the relation between gender and mathematical performance, mediation analyses were performed. Factors were entered as possible mediators when the regression analysis showed that there was a unique contribution to the mathematical performance test, and when the *t*-tests indicated that gender differences were present. A (bootstrapped) mediation analysis shows whether there is a significant indirect effect (quantified as the product of the unstandardized path coefficients, *ab*) of the mediator(s) that account for some portion of the total effect (specified as *c*) between the predictor and outcome variable. The remaining (unmediated) direct effect of the predictor on the outcome variable is specified as *c'*. Here, the model is restricted by the assumption that $c = ab + c'$. Contrary to a standard multiple regression analysis, this analysis explicitly tests to which extent the mediating variable can explain the relation between gender and mathematical performance. There is a full mediation, if *ab* is significant but *c* is not, while there is a partial mediation when both *ab* and *c* are significant (Preacher & Hayes, 2008). The conclusions about statistical validity are based on the confidence intervals: in case the confidence intervals do not contain zero, the effect is significant. In addition to these pre-registered main analyses, we also pre-registered a hierarchical regression analysis to examine which cognitive factors contribute to arithmetic performance and cognitive reflection which can be found on the Open Science Framework: https://osf.io/8vwe7

3. Results

3.1. Relations with mathematical performance

Firstly, it was examined whether working memory, math anxiety and gender-related stereotypes explain unique variance in mathematical performance as measured by an arithmetic fluency test and cognitive reflection test. Table 2 displays the bivariate correlations between the measures. A first regression analyses with the TTR as outcome measure and working memory, math anxiety and gender-related stereotypes as predictors showed that arithmetic fluency was predicted by performance on the Digit Span Forward task, Math Anxiety and Gender-related stereotypes about masculine items (see Table 4). A second regression analyses with the CRT as outcome measure and working memory, math anxiety and gender-related stereotypes as predictors revealed that that only Math Anxiety was a significant predictor (see Table 5).

3.2. Accounting for the relation between gender differences and mathematical performance

To examine which factors could account for the relation between gender and mathematical performance, mediation analyses were conducted. Factors were entered as possible mediators when the regression analyses showed that there was a unique contribution to the mathematical performance tests and when *t*-test revealed gender differences. Bootstrapped mediation analyses were conducted with the Preacher and Hayes (2008) SPSS MEDIATE' macro. In these mediation analyses, gender was the predictor variable and mathematical performance as measured by the two tests the outcome variable.

Fig. 1 presents the mediation analysis results with gender as predictor and arithmetic as outcome variable. Math anxiety and explicit gender-related stereotypes were entered as mediators because of the unique contribution to arithmetic and because gender differences were present for these factors (see Table 3). Results revealed that the relation between gender and arithmetic was partially mediated by math anxiety, $ab = -5.087$, $SE = 1.812$, $CI = -9.153$ to -2.072 and explicit gender-related stereotypes for masculine items, $ab = -2.266$, $SE = 1.157$ $CI = -4.917$ to -0.406.

Table 2
Bivariate correlations between the measures.

		1	2	3	4	5	6	7	8	9	10	11
1	Gender											
2	Arithmetic											
3	Cognitive reflection		0.31**									
4	Number series completion		0.36**	0.27**								
5	PL		0.29**	0.13	0.16*							
6	PL * CE		0.26**	0.14	0.14	0.55**						
7	VSSP		0.34**	0.12	0.25**	0.18*	0.32**					
8	VSSP * CE		0.25**	0.16*	0.33**	0.04	0.29**	0.59**				
9	Math anxiety		−0.39**	−0.22**	−0.23**	−0.17	−0.06	−0.18*	−0.19**			
10	Explicit stereotypes (masculine)		0.29**	−0.06	0.13	0.06	0.11	0.10	0.11	−0.09		
11	Explicit stereotypes (feminine)		0.22**	−0.07	0.10	0.06	0.05	0.05	0.08	−0.05	0.87**	
12	Implicit stereotypes		0.02	−0.01	0.04	0.02	0.09	0.13	0.08	0.06	0.16*	0.17*

Note. PL = phonological loop, VSSP = visuo-spatial sketchpad, CE = central executive.
* $p < .05$.
** $p < .01$.

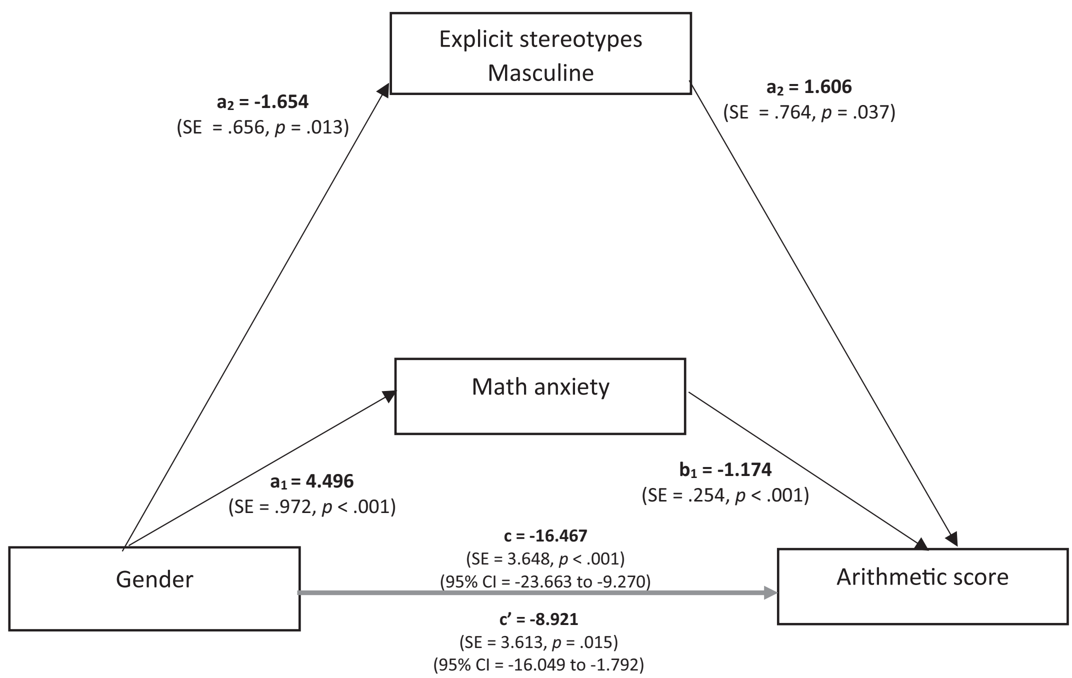

Fig. 1. Mediation model with arithmetic score as outcome variable, gender as predictor and explicit stereotypes and math anxiety as mediators.

Table 3
Bivariate correlations between the measures disaggregated by gender.

		1	2	3	4	5	6	7	8	9	10	11
1	Arithmetic	–	0.29**	0.32**	0.27**	0.27**	0.31**	0.13	−0.43**	0.21*	−0.20*	0.16
2	Cognitive reflection	0.24**	–	0.40**	0.12	0.14	0.10	0.13	−0.22*	−0.17	−0.14	0.09
3	Number series completion	0.35**	0.07	–	0.12	0.02	0.22*	0.30**	−0.22*	0.07	0.10	−0.09
4	PL	0.35**	0.15	0.23*	–	0.58**	0.04	−0.17	−0.21*	−0.08	−0.08	0.06
5	PL * CE	0.32**	0.17	0.30**	0.52**	–	0.18	0.13	−0.08	0.11	0.07	−0.01
6	VSSP	0.32**	0.10	0.26*	0.30**	0.44**	–	0.48**	−0.21	0.10	0.10	−0.06
7	VSSP * CE	0.34**	0.16	0.35**	0.24*	0.45**	0.68	–	−0.25	0.16	0.18	−0.05
8	Math anxiety	−0.17	−0.11	−0.14	−0.15	−0.10**	−0.09	−0.08	–	−0.04	−0.06	0.02
9	Explicit stereotypes (masculine)	0.28*	−0.02	0.13	0.17	0.13	0.07	0.05	−0.03	–	0.89**	0.06
10	Explicit stereotypes (feminine)	0.19	−0.05	0.06	0.20	0.05	−0.03	−0.05	0.05	0.86**	–	0.07
11	Implicit stereotypes	−0.06	−0.13	0.05	0.19	0.12	0.26*	0.21*	−0.05	0.18	0.17	–

Note. The results for the female sample are shown above the diagonal. The results for the male sample are shown below the diagonal.
* $p < .05$.
** $p < .01$.

When conducting an exploratory reversed mediation with arithmetic as mediator and math anxiety as dependent variable, results showed that arithmetic mediated the relation between gender and math anxiety, *ab* = 1.413, *SE* = 0.496, *CI* = 0.605 to 2.494 suggesting a mutual influence between math anxiety and arithmetic performance. These results can be found on the Open Science Framework: https://osf.io/8vwe7

Table 4
Linear regression with arithmetic as dependent variable.

Independent variables	Standardized β	t	p	R^2
PL	0.17	2.27	0.024*	0.33**
PL * CE	0.05	0.66	0.513	
VSSP	0.13	1.70	0.090	
VSSP * CE	0.11	1.44	0.151	
Math anxiety	−0.31	−4.79	<0.001	
Gender-related stereotypes masculine	0.26	2.06	0.041	
Gender-related stereotypes feminine	−0.04	−0.33	0.743	

* $p < .05$.
** $p < .01$.

Table 5
Linear regression with cognitive reflection as dependent variable.

Independent variables	Standardized β	t	p	R^2
PL	0.05	0.53	0.600*	0.30**
PL * CE	0.09	0.93	0.355	
VSSP	−0.06	−0.63	0.530	
VSSP * CE	0.15	1.59	0.113	
Math anxiety	−0.22	−2.85	0.005	
Gender-related stereotypes masculine	−0.07	−0.48	0.632	
Gender-related stereotypes feminine	−0.02	−0.33	0.922	

* $p < .05$.
** $p < .01$.

Fig. 2 presents the mediation analysis results with gender as predictor, cognitive reflection as outcome variable, and math anxiety as mediator. Math anxiety was entered as a mediator because this factor differed between men and women and because it was significantly correlated with cognitive reflection (see Table 2). Results showed that math anxiety was a significant mediator of the relation between gender and cognitive reflection. More specifically, results demonstrated a significant total effect and a significant effect of the mediator variable math anxiety, $ab = -0.428$, $SE = 0.266$, $CI = 0.032$ to 1.059. This concerned a full mediation since the direct effect of gender on cognitive reflection was no longer significant, indicating that math anxiety fully accounted for the relationship between gender and cognitive reflection.

An exploratory reversed mediation with cognitive reflection as mediator and math anxiety as dependent variable, revealed that cognitive reflection mediated the relation between gender and math anxiety, $ab = 0.428$, $SE = 0.266$, $CI = 0.032$ to 1.059. However, the direct path was still significant. These results suggest a mutual influence, but also demonstrate that math anxiety influences cognitive reflection more than vice versa. These results can be found on the Open Science Framework: https://osf.io/8vwe7

In conclusion, the mediation analyses showed that the relation between gender and arithmetic performance was partially mediated by math anxiety and explicit gender-related stereotypes for masculine items. In addition, gender differences in cognitive reflection were fully mediated by math anxiety. When interpreting these results, it is important to note that the relation between gender and arithmetic was stronger than the relation between gender and cognitive reflection. As a consequence, when taking math anxiety into account as mediator, the relation between gender and arithmetic was reduced to a larger extent than the relation between gender and cognitive reflection.

4. Discussion

In a world where women are still underrepresented in the STEM field (UNESCO, 2017), it can be questioned whether there are gender differences in mathematical performance and if so, which factors contribute to this. While previous studies have examined gender differences in mathematical performance, these studies have often focused on one specific mathematical test and on one specific factor that could possibly underlie gender differences in mathematical performance. The current study attempted to fill these gaps in the literature by examining whether working memory, math anxiety and gender-related stereotypes could account for gender differences in mathematical performance on a basic and complex mathematical test.

Results showed gender differences for the arithmetic fluency test and cognitive reflection test. These findings reveal that men and women differ in mathematical performance for both basic and complex mathematical tests, a finding which has also been consistently demonstrated in previous research (Cueva et al., 2016; Else-Quest et al., 2010; Lindberg et al., 2010; Lynn & Irwing, 2008; Miller & Bichsel, 2004; Primi et al., 2018). Subsequently, results revealed that arithmetic fluency was predicted by performance on the Digit Span Forward task, math anxiety and explicit gender-related stereotypes about masculine items and that gender differences were present for math anxiety and explicit gender-related stereotypes. No relation was observed between arithmetic and implicit gender-related stereotypes. However, it should be acknowledged that the questionnaire measuring explicit gender stereotypes used in the current study included many general items (e.g., ride a motorcycle). The questionnaire did not assess explicit gender-related stereotypes that were directly related to math. Nevertheless, results revealed a relation with arithmetic performance and therefore it was further examined whether math anxiety and explicit gender-related stereotypes could account for the observed gender differences in arithmetic performance.

In order to test the hypothesis that math anxiety and gender-related stereotypes mediate the relation between gender and mathematical performance, mediation analyses were performed. Results showed that the relation between gender and arithmetic performance was partially mediated by math anxiety and explicit gender-related stereotypes. This finding is line with the results of Rossi et al. (2021), also revealing a

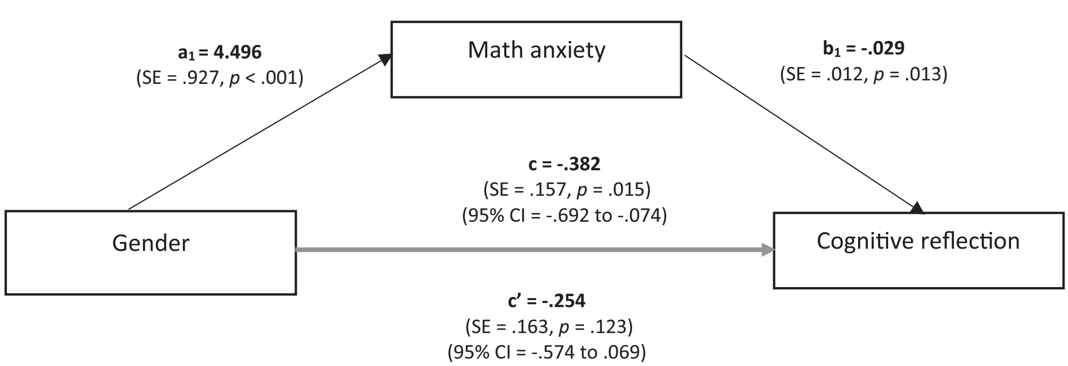

Fig. 2. Mediation model with cognitive reflection as outcome variable, gender as predictor and math anxiety as mediator.

relationship between gender-related stereotypes, math anxiety and arithmetic performance indicating that the endorsement of mathematics gender stereotypes can lead to increased math anxiety, which affects arithmetic performance. In addition, the relation between gender and cognitive reflection was fully mediated by math anxiety. A similar finding has been observed by Primi et al. (2018), indicating an indirect effect of gender on the CRT through math anxiety.

When interpreting the results of the cognitive reflection test, a few issues need to be taken into account. A first issue is that while some researchers assume that the cognitive reflection test merely measures mathematical skills (Sinayev & Peters, 2015), others have suggested that the CRT also measures reasoning skills (Pennycook & Ross, 2016). Therefore, it could be argued that the gender differences in cognitive reflection reflect gender differences in reasoning skills. However, previous studies suggest that gender differences observed on the CRT can probably be attributed to the numerical nature of the task as no gender differences have been found for a verbal CRT (Sirota et al., 2021) or when numerical skills and math anxiety were taken into account (Primi et al., 2018). Therefore, it is likely that the gender differences in cognitive reflection in the current study are a result of gender differences in mathematical skills and not a result of gender differences in reasoning skills. In the current study, we focused on the CRT because it not only involves mathematical word problems but also taps into the ability to reflect on the problem structure and strategy. By using the CRT, we aimed to avoid that participants could solve the problem with a shallow approach based on textual cues. Nevertheless, research has also shown gender differences for other forms of mathematical word problems. For example, Reinhold et al. (2020) observed gender differences in adults for mathematical word problems. Interestingly, these gender differences could not be fully explained by spatial ability, verbal ability, numerical ability and general reasoning. On the basis of these results, Reinhold et al. (2020) suggested that gender-specific attitudes, beliefs and emotions probably underlie gender differences in mathematical word problems.

A second issue is that the reliability of the CRT in the current study was relatively low, although this is in line with previous research showing a range between 0.60 and 0.74 (Liberali et al., 2012 Weller et al., 2013; for a review, see Campitelli & Gerrans, 2014). This low internal consistency is mainly due to the small number of items. Therefore, it might be helpful for future studies to use cognitive reflection tests with more items and a higher internal consistency such as the 7-item cognitive reflection test (Ring et al., 2016). Despite the low internal consistency of the CRT, it is important to note that performance on the cognitive reflection test is stable over time (Stagnaro et al., 2018) and that the CRT correlates with other measures of cognitive ability (Otero et al., 2022) indicating that the test has a sufficient reliability and validity.

By including both an arithmetic fluency test and cognitive reflection test, we could examine the contribution of working memory, math anxiety and gender-related stereotypes for gender differences in basic and complex mathematical tests. Arithmetic performance was related to both gender-related stereotypes and math anxiety. Cognitive reflection was only related to math anxiety. It can be questioned why explicit gender-related stereotypes about masculine items are associated with arithmetic performance but not with cognitive reflection. Possibly, arithmetic tasks more obviously tap into numerical skills than cognitive reflection. As a consequence, arithmetic is perceived as a task on which men perform better and might therefore correlate more strongly with explicit gender related stereotypes about men than cognitive reflection. For both tests, gender differences were observed and these gender differences were largely explained by math anxiety. Math anxiety only partially mediated the relation between gender and arithmetic fluency while math anxiety fully mediated the relation between gender and cognitive reflection. The finding that math anxiety was a full mediator in the relation between gender and arithmetic but only a partial mediator in the relation between gender and cognitive reflection is in line with our expectations. However, this finding can be explained by the stronger relation between gender and arithmetic compared to the relation between gender and cognitive reflection. Consequently, the relation between gender and arithmetic was reduced to a larger extent than the relation between gender and cognitive reflection when math anxiety was taken into account as mediator.

Altogether, the current study's findings show that gender differences in mathematical performance is strongly related to math anxiety. Since math anxiety is a mediator in the relation between gender and mathematical performance, this raises the question of what drives gender differences in math anxiety and, consequently the differences in mathematical performance. To answer this question, we contemplate the theories about the relation between math anxiety and mathematics. The most widely held theory explaining the relation between anxiety and cognitive tasks states that when confronted with a cognitive task (such as math), negative thoughts and ruminations compete with the ongoing cognitive task for working memory resources. A consequence of this competition is lower cognitive efficiency, resulting in increased reaction time or decreased accuracy on the cognitive task. This theory, first introduced by Eysenck and Calvo (1992) as the processing efficiency theory, was later extended by Ashcraft and Kirk (2001) to math anxiety. In line with this theory, a large body of research has consistently shown that high math anxious individuals must exert greater cognitive effort to achieve the same level of performance attained by low math anxious individuals (Ashcraft & Kirk, 2001). In sum, the processing efficiency theory states that individual differences in math anxiety affect performance in mathematics because of a disruption of working memory.

While we did not directly assess the occupation of working memory during the mathematical tests, the current study was introduced as a study examining mathematical skills, and mathematical tests were alternated with working memory tests. Based on the processing efficiency theory and previous findings showing lower working memory performance among high-math anxious individuals (Ashcraft & Kirk, 2001), one would expect that since women reported more math anxiety, they would also perform worse on the working memory tests. However, we did not observe gender differences in working memory. Thus, the processing efficiency theory could not explain gender differences in math anxiety and mathematical ability in the current study.

Rather, our findings can be integrated within the interpretation framework of Ramirez et al. (2018). According to this framework, math anxiety is dependent on how individuals interpret their math-related experiences. The interpretation framework further assumes that math anxiety is largely determined by how individuals appraise previous math experiences rather than the outcomes themselves. Thus, the beliefs about math (rather than actual working memory capacity) may explain the observed gender differences in math anxiety and mathematical performance. Unfortunately, the current study's design does not allow to test women's beliefs about math directly, since the AMAS considers only very few aspects of math anxiety, namely learning math anxiety and math evaluation anxiety (Cipora et al., 2019). The AMAS does not test everyday math anxiety or general attitudes towards mathematics. However, in line with the interpretation framework, we found that men and women differ concerning both learning math anxiety and math evaluation anxiety. Thus, women are not only more anxious to be evaluated for math, but they are also more anxious to engage in math-related activities, indicating that they appraise mathematics differently. It is important to acknowledge that the hypothesis that appraisal processes underlie gender differences in math anxiety and mathematical performance could not be explicitly tested in the current cross-sectional study. Besides the significant mediation of math anxiety in the relation between gender and mathematical performance, reversed mediation models showed that mathematical performance was also a significant mediator in the relation between gender and math anxiety. Previous studies obtained similar results, suggesting a bidirectional relationship between math anxiety and mathematical performance (Carey et al., 2016). Longitudinal research should identify the direction of the relation and factors that underlie gender differences in reported math anxiety. In future research, it could also be informative to conduct a path analysis which includes all the variables in one model. The sample size of the current study was too small to infer reliable results from a path analysis (Wang & Rhemtulla, 2021: Wolf et al., 2013). However, future studies with a larger sample size could examine multiple potential mediators within one model to see how the variables interact.

Altogether, the current study showed that there are gender differences among adults in arithmetic and cognitive reflection. Math anxiety appeared to be a significant mediator in the relation between gender and mathematical performance. Probably this is because math anxiety primarily reflects how individuals appraise previous math experiences rather than cognitive differences and as a consequence, the appraisal of previous math experiences affects mathematical performance.

Statement regarding concordance with human subject guidelines

The research reported in this article involves healthy human participants and does not utilise any invasive techniques, substance administration or psychological manipulations. Therefore, this study only required, and received, approval from the internal Ethics Committee. Furthermore, this research was conducted, and written informed consent of each participant was obtained, according to the principles expressed in the Declaration of Helsinki. All the participants in the experiment had previously provided written consent when signing up online to participate in this experiment. In doing so, they had indicated to have read and to have agreed with both the rules regarding participation and the researchers' commitments and privacy policy. They are also informed that they can stop participating in the experiment whenever they want to do so. All data were analysed anonymously.

Declarations of competing interest

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

References

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49*, 5–28. https://doi.org/10.1080/713755608

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158. https://doi.org/10.1037/0033-295x.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation, 8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., & De Smedt, B. (2019). Gender equality in four- to five-year-old preschoolers' early numerical competencies. with De Smedt, B *Developmental Science, 22*, Article e12718. https://doi.org/10.1111/desc.12718.

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*, 134–168. https://doi.org/10.1037/bul0000307

Barton, A., Matthews, B., Farmer, E., & Belyavin, A. (1995). Revealing the basic properties of the visuospatial sketchpad: The use of complete spatial arrays. *Acta Psychologica, 89*, 197–216. https://doi.org/10.1016/0001-6918(95)98944-Y

Beilock, S., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology, 136*, 256–276. https://doi.org/10.1037/0096-3445.136.2.256

Beilock, S. L., & Maloney, E. A. (2015). Math anxiety: A factor in math achievement not to be ignored. *Policy Insights from the Behavioural and Brain Sciences, 2*, 4–12. https://doi.org/10.1177/2372732215601438

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*, 434–447. https://doi.org/10.3758/s13421-013-0367-9

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology, 6*, 1987. https://doi.org/10.3389/fpsyg.2015.01987

Caviola, S., Toffalini, E., Giofrè, D., Mercader Ruiz, J., Szűcs, D., & Mammarella, I. (2022). Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants. *Educational Psychology Review, 34*, 363. https://doi.org/10.1007/s10648-021-09618-5

Cipora, K., Artemenko, C., & Nuerk, H.-C. (2019). Different ways to measure math anxiety. In I. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety: What is known, and what is still missing* (1st ed., pp. 20–41). Routledge.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Perez, E., Ponti, G., Sartarelli, M., Yu, H., & Zhukova, V. (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics, 64*, 81–93. https://doi.org/10.1016/j.socec.2015.09.002

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*, 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

De Vos, T. (1992). *Tempo-test-rekenen, T.T.R.: Test voor het vaststellen van het rekenvaardigheidsniveau der elementaire bewerkingen (automatisering) voor het basis- en voortgezet onderwijs*. Lisse, Nederland: Swets & Zetilinger.

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386. https://doi.org/10.1080/09541440244000328

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. https://doi.org/10.1037/a0018053

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23. https://doi.org/10.1111/1467-8721.00160

Ertl, B., Luttenberger, S., & Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology, 8*, 703. https://doi.org/10.3389/fpsyg.2017.00703

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409–434.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*, 336–353.

Finell, J., Sammallahti, E., Korhonen, J., Eklöf, H., & Jonsson, B. (2022). Working memory and its mediating role on the relationship of math anxiety and math performance: A meta-analysis. *Frontiers in Psychology, 6643*. https://doi.org/10.3389/fpsyg.2021.798090

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. https://doi.org/10.1016/j.jsp.2014.10.002

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42. https://doi.org/10.1257/089533005775196732

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. Reis, & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York, NY, USA: Cambridge University Press.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Hart, S. A., & Ganley, C. M. (2020). The nature of math anxiety in adults: Prevalence and correlates. *Journal of Numerical Cognition, 5*, 122–139. https://doi.org/10.5964/jnc.v5i2.195

Hickendorff, M. (2021). The demans of simple and complex arithmetic word problems on language and cognitive resources. *Frontiers in Psychology, 12*, 1–12.

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences, 48*, 45–53. https://doi.org/10.1016/j.lindif.2016.02.006

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. A. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10*, 178–182. https://doi.org/10.1177/1073191103010002008

Hutchison, J., Lyons, I., & Ansari, D. (2019). More similar than different: Gender differences in basic numeracy are the exception not the rule. *Child Development*, e66–e79. https://doi.org/10.1111/cdev.13044

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494–495. https://doi.org/10.1126/science.1160364

Jansen, B. R., Schmitz, E. A., & van der Maas, H. L. (2016). Affective and motivational factors mediate the relation between math skills and use of math in everyday life. *Frontiers Psychology, 7*, 513. https://doi.org/10.3389/fpsyg.2016. 00513

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitivereflection predicts real-life decision outcomes, but not over and abovepersonality and decision-making styles. *Journal of Behavioral Decision Making, 29*, 52–59. https://doi.org/10.1002/bdm.1875

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35*, 211–223. https://doi.org/10.1016/j.intell.2006.07.009

Lachance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*, 195–216. https://doi.org/10.1016/j.lindif.2005.12.001

Liben, L. S., Bigler, R. S., Ruble, D. N., Martin, C. L., & Powlishta, K. K. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67*, 1–183. https://doi.org/10.1111/1540-5834.t01-1-00187

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. *Journal of Behavioral Decision Making, 25*, 361–381. https://doi.org/10.1002/bdm.752

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135. https://doi.org/10.1037/a0021276

Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management, 11*, 311. https://doi.org/10.2147/PRBM.S141421

Lynn, R., & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g defined as working memory capacity. *Intelligence, 36*, 226–235. https://doi.org/10.1016/j.intell.2007.06.002

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*, 591–606.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459–496.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., & Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*, 10593–10597. https://doi.org/10.1073/pnas.0809921106

Organization for Economic Cooperation and Development. (2016). *PISA 2015 results in focus*. OECD Publishing.

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence, 90*, Article 101614. https://doi.org/10.1016/j.intell.2021.101614

Pauls, F., Petermann, F., & Lepach, A. C. (2013). Gender differences in episodic memory and visual working memory including the effects of age. *Memory, 21*, 857–874. https://doi.org/10.1080/09658211.2013.765892

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*, 455–473. https://doi.org/10.1037/edu0000079

Pennycook, G., & Ross, R. M. (2016). Commentary on: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology, 7*, 9. https://doi.org/10.3389/fpsyg.2015.00532

Piccardi, L., D'Antuono, G., Marin, D., Boccia, M., Ciurli, P., Incoccia, C.Guariglia, C., … (2019). New evidence for gender differences in performing the corsi test but not the digit span: Data from 208 individuals. *Psychological Studies, 64*, 411–419. https://doi.org/10.1007/s12646-019-00512-3

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891. https://doi.org/10.3758/BRM.40.3.879

Price, G. R., Mazzocco, M. M., & Ansari, D. (2013). Why mental arithmetic counts: Brain activation during single digit arithmetic predicts high school math scores. *Journal of Neuroscience, 33*, 156–163. https://doi.org/10.1523/jneurosci.2936-12.2013

Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking and Reasoning, 24*, 258–279. https://doi.org/10.1080/13546783.2017.1387606

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist, 53*, 145–164. https://doi.org/10.1080/00461520.2018.1447384

Reinhold, F., Hofer, S., Berkowitz, M., Strohmaier, A., Scheuerer, S., Loch, F., et al. (2020). The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults. *Mathematics Education Research Journal, 32*, 189–211. https://doi.org/10.1007/s13394-020-00331-0

Ring, P., Neyse, L., David-Barett, T., & Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Frontiers in Psychology, 7*, 1680. https://doi.org/10.3389/fpsyg.2016.01680

Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology, 18*(3), 378–397. https://doi.org/10.1080/09541440500234104

Rossi, S., Xenidou-Dervou, I., Simsek, E., Artemenko, C., Daroczy, G., Nuerk, H., & Cipora, K. (2021). *Mathematics-gender stereotype endorsement influences mathematics anxiety, self-concept, and performance differently in men and women*. https://doi.org/10.31234/osf.io/f52zg. September 20.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology, 6*, 532. https://doi.org/10.3389/fpsyg.2015.00532

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., & Marshall, A. C. (2021). Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making, 34*, 322–343. https://doi.org/10.1002/bdm.2213

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision making, 13*, 260–267. https://doi.org/10.2139/ssrn.3115809

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles, 64*, 324–335. https://doi.org/10.1007/s11199-010-9924-x

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cognitive Affective Behavioral Neuroscience, 16*, 3–22. https://doi.org/10.3758/s13415-015-0370-7

Toplak, M. E., West, R. F., & Stanovich, K. E. (2017). Real-world correlates of performance on heuristics and biases tasks in a community sample. *Journal of Behavioral Decision Making, 30*, 541–554. https://doi.org/10.1002/bdm.1973

UNESCO. (2017). Cracking the code: girls' and women's education in science, technology, engineering and mathematics. Retrieved November 2, 2020, from https://unesdoc.unesco.org/ark:/48223/pf0000253479.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM, 52*, 1–16. https://doi.org/10.1007/s11858-020-01130-4

Vuletich, H. A., Kurtz-Costes, B., Cooley, E., & Payne, B. K. (2020). Math and language gender stereotypes: Age and gender differences in implicit biases and explicit beliefs. *PLoS One, 15*, Article e0238230. https://doi.org/10.1371/journal.pone.0238230

Wang, M.-T., & Degol, J. L. (2016). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*, 119–140. https://doi.org/10.1007/s10648-015-9355-x

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1). https://doi.org/10.1177/2515245920918253

Wechsler, D. (2012). *Wechsler Adult Intelligence Scale - Fourth Edition - Nederlandse bewerking. Afname- en scoringshandleiding*. Amsterdam, Nederland: Pearson Assessment and Information BV.

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. *Journal of Behavioral Decision Making, 26*, 198–212. https://doi.org/10.1002/bdm.1751

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*, 913–934. https://doi.org/10.1177/0013164413495237

Zilles, D., Lewandowski, M., Vieker, H., Henseler, I., Diekhof, E., Melcher, T.Gruber, O., … (2016). Gender differences in verbal and visuospatial working memory performance and networks. *Neuropsychobiology, 73*, 52–63. https://doi.org/10.1159/000443174