Gender differences in mathematics anxiety: implicit and explicit measurements
Heemskerk, Jasmijn

Gender differences in mathematics anxiety:

Implicit and explicit measurements

Masterthesis

Developmental Psychopathology in Education and Child Studies (research)

Jasmijn Heemskerk, s1503340

27-10-2021

Supervisor: Dr. T.M. Sikkema-de Jong

Second reader: Prof. dr. C.A. Espin

Abstract

The gender gap in STEM majors and careers is very prominent and alarming (Beilock & Maloney, 2015). Students are more likely to avoid STEM careers when they show higher levels of Mathematics Anxiety (MA; Ahmed, 2018). The current study investigates the difference in Mathematics Anxiety between male and female students in their first year of the preacademic educational track (HAVO, atheneum and gymnasium) in secondary school in the Netherlands. The Numerical Dot-Probe Task (NDPT) is a computerized task that is based on attentional bias and less susceptible to bias than self-report questionnaires. (Rubinsten, Eidlin, Wohl, & Akibli, 2015). Related concepts to MA that will be controlled for are; mathematical achievement, general anxiety and working memory.

Participants were recruited using convenience sampling, the final sample consisted of seven students. Using the adapted t-test, z-scores of all participants were compared in order to answer the research question (Crawford, Garthwaite, & Wood, 2010). 100% of girls and 20% of boys showed higher levels of MA when looking at their scores on the self-report questionnaires compared to their score on the NDPT. Overall, the girls' MA score was overestimated and the boys' MA score was underestimated. However, this effect was not strong enough to result in an overrepresentation of girls and an underrepresentation of boys in the population of children with high levels of MA when looking at the self-report questionnaires. Results suggested that the prominent gender gap in STEM studies might not be due to gender differences in MA. Due to the smaller sample size, it is important that these findings are replicated in future studies using bigger samples. Stereotype threat is a variable that could explain the STEM gender gap and should thus be included in these future studies.

Keywords: gender differences, Numerical Dot-Probe Task, mathematics anxiety, mathematical achievement, general anxiety, working memory, preacademic educational track, secondary school

Introduction and theoretical background

According to recent data collected by the main statistical office in the Netherlands, 'Het Centraal Bureau voor de Statistiek' [The Central Bureau of Statistics] (CBS, 2020), there are less female students enrolled in bèta studies/majors (science, technology, engineering and mathematics; STEM) compared to male students (CBS, 2020). The gender gap remains prominent in STEM classrooms and work environments even though there has been a small increase of female students enrolled in STEM majors due to policy changes and attention from the government (Beilock & Maloney, 2015). The gender gap in STEM majors and STEM careers is alarming, it contributes to the gender wage gap in favour of male employees in the labor markets since job positions in STEM careers typically pay more (Brown & Corcoran, 1997; Black, Haviland, Sanders, & Taylor, 2008; Blau & Kahn, 2017). The gender gap results in a loss of talented employees for STEM job positions where there is already a employee shortage (Aguinis, Ji, & Joo, 2018; McDonald & Waite, 2019).

It is important to understand the mechanisms that underlie this gender gap in STEM majors and consequently STEM careers to provide equal opportunities for both male and female students in these fields. Longitudinal studies have showed that an interest in STEM classes in high school predicts STEM career choices in adulthood (Benbow, 2012; Wang, Eccles, & Kenny, 2013). Interventions thus typically have better outcomes when executed during the first year of secondary school; before students make up their minds regarding STEM majors and STEM careers (Ahmed, 2018). An important finding has been that these students are more likely to avoid STEM careers when they show higher levels of Mathematics Anxiety (MA; Ahmed, 2018; Moakler & Kim, 2014). The current study will explore the possible gender gap in MA that can possibly explain the gender gap in STEM careers.

Mathematics Anxiety

It has been found that MA is related to general anxiety but that they are not the same construct since they have only 37% of shared variance (Hembree, 1990). MA is an adverse emotional reaction to (the anticipation of) mathematics or mathematical situations (Ashcraft & Ridley, 2005). A definition that is often used in both research and practice regarding MA is; “a feeling of tension, apprehension, or even dread that interferes with the ordinary manipulation of numbers and the solving of mathematical problems” (Ashcraft & Faust, 1994, p.98). MA is not included in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V; American Psychiatric Association, 2013). However, MA does show the characteristics of a specific phobia: it is a state anxiety reaction, there is an elevated cognitive and physiological arousal and it is a specific fear for a stimulus or situation (Ashcraft & Ridley, 2005). MA is an affective factor that negatively effects mathematical learning performance and the student’s confidence in their mathematical abilities (Beilock & Maloney, 2015). Higher levels of MA in a student have been associated with lower confidence in their mathematical abilities (Ahmed, Minnaerd, Kuyper, & van der Werf, 2012). Individuals with higher levels of MA become nervous when engaging in mathematical tasks and naturally avoid mathematics and mathematics-related professions which limits their future opportunities in STEM careers (Ashcraft & Ridley, 2005; Hembree, 1990; Suárez-Pellicioni, Núñez-Peña, & Colomé, 2016). More recently, Ahmed (2018) found that students that experienced high levels of MA in adolescence were more likely to avoid STEM majors and STEM careers compared to their classmates.

Mathematical performance. Since individuals that experience MA become nervous when engaging in mathematical tasks and make an effort to avoid these tasks, the link between MA and mathematical performance has been investigated by many researchers (Hembree, 1990; Ma, 1999). A study among college freshmen found that even in its milder forms, MA is negatively related to mathematical performance and participation in a STEM major or career (Moakler & Kim, 2014). When including multiple studies with a range of age groups in a meta-analysis, from children to adults, a moderate negative relationship between MA and mathematical performance is found (Hembree, 1990; Ma, 1999). This relationship is bidirectional since early mathematics performance negatively affects the level of MA in a student, which in turn negatively affects later mathematics performance (Lou et al., 2014).

Working memory. Working memory has been pointed out as a possible mediator in the negative relationship between MA and mathematical performance (Namkung, Peng, & Lin, 2019). According to the processing efficiency theory, anxiety involves worrying intrusive thoughts that occupy the limited resources of the central executive of the working memory (Eysenck & Calvo, 1992). The inhibition theory states that individuals with higher levels of anxiety also have difficulty inhibiting attention to anxious thoughts (Ashcraft, Kirk, & Hopko, 1998). To combine the processing efficiency theory and the inhibition theory: tasks that demand much of the central executive's resources or tasks have anxiety evoking properties are more effected by anxiety (Ashcraft & Moore, 2009). Working memory has a mediating role in the negative relationship between MA and mathematical performance (Hopko, Ashcraft, Gute, Ruggiero, & Lewis, 1998). If a student has a good working memory, the negative relationship between MA and mathematical performance will decrease. There also is a direct negative relationship between both visual and verbal working memory and MA; when a student has higher levels of MA, they show a smaller working memory span (Ashcraft & Kirk, 2001; Miller & Bichsel, 2004).

Gender. Across many grades, from kindergarten through undergraduate, female students report higher levels of MA compared to male students when using self-report questionnaires to measure the students' level of MA (Beilock & Maloney, 2015; Hembree, 1990). One hypothesis that might explain the gender gap in STEM majors and STEM careers is the fact that MA is more prevalent in females than in males which might make female students more likely to avoid mathematics and mathematics related professions (Ahmed, 2018; Beilock & Maloney, 2015; Hembree, 1990). Two meta-analyses showed that studies investigating MA have predominantly used self-report questionnaires to assess a student's level of MA (Hembree, 1990; Ma, 1999). The use of self-report questionnaires has been debated since these are susceptible to different biases that can influence the reliability and validity of the questionnaire (Huang, Liao, & Chang, 1998). Responses on a self-report questionnaire are influenced by biases like social desirability and the relationships among the variables can be enhanced, suppressed or even produced (King & Bruner, 2000). The earlier mentioned hypothesis that used MA as a possible explanation for the gender gap in STEM careers can be questioned since the most popular method of assessing MA is susceptible to biases that might influence the proportion of male and female students reported to have high levels of MA (van de Mortel, 2008).

Some factors inherent in the child might result in an overrepresentation of female students and an underrepresentation of male students. Male students are more likely to underreport fear (Pierce & Kirkpatrick, 1992) and anxiety (Egloff & Schmukle, 2004) and overestimate their mathematical performance (Bench, Lench, Liew, Miner, & Flores, 2015). On top of that, females are more willing to admit their anxiety and they are more self-critical of their performance compared to their male counterparts (Ashcraft, 2002). Next to child factors, some societal factors can influence the students' responses on self-report questionnaires (van de Mortel, 2008). The stereotype that mathematics is for boys can be found in both boys and girls as early as elementary school and in many different nations (Cvencek, Meltzoff, & Greenwald, 2011; Nosek et al., 2009). This stereotype influences the way students answer questions in a self-report questionnaire through the social desirability response bias (van de Mortel, 2008). On top of that, the mathematical self-concept of male students can be enhanced due to their perception of gender stereotypes on mathematics causing them to answer the questions more positively (Bench et al., 2015; Kurtz-Costes, Rowley, Harris-Britt, & Woods, 2008). This might imply that there is no difference in levels of MA between male and female students but that the assessment method underreports male students and overreports female students. The current study aims to investigate the difference in MA between male and female students in their first year of the preacademic educational track (HAVO, atheneum and gymnasium) in secondary school in the Netherlands using different measurement methods.

Attentional Bias

The use of more implicit measures that use these cognitive processes to assess anxiety might result in a more objective measure of MA compared to the often-used self-report questionnaires that are susceptible to bias. An attribute associated with anxiety disorders in general and MA in specific is attentional bias (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Suárez-Pellicioni, Núñez-Peña, & Colomé, 2015). This bias can be described as the differential allocation of attention toward the threatening stimulus compared to the neutral, non-threatening, stimulus (Bar-Haim et al., 2007). It is still debated whether individuals with higher levels of anxiety are faster to engage with an anxiety provoking stimulus (vigilance) or whether they have difficulty disengaging from this anxiety provoking stimulus (maintenance) (Bar-Haim et al., 2007). Attentional bias can utilized in the development of implicit measurements for anxiety in an individual (Bar-Haim et al., 2007).

Initially, the emotional Stroop task was the preferred measurement for the attentional bias, but it suffered much criticism (Bar-Haim et al., 2007; MacLeod, Mathews, & Tata, 1986). To overcome these criticisms, the dot-probe paradigm was developed (MacLeod et al., 1986). The Dot-Probe Task takes advantage of the attentional bias toward anxiety evoking information to assess anxiety levels (Miyazaki, Ichihara, Wake, & Wake, 2012; Sass et al., 2010). To give an appropriate indication of the individual's level of MA, the standard Dot-Probe Task has been adapted with mathematical stimuli as the anxiety evoking information (Rubinsten, Eidlin, Wohl, & Akibli, 2015). This new, adapted, version of the Dot-Probe Task is called the Numerical Dot-Probe Task and will be used in the current study as an implicit and possibly more reliable measurement of MA.

The Numerical Dot-Probe Task is a computerized task that uses the differences in reaction time for anxiety evoking and neutral stimuli as a measure for MA that is based on attentional bias (Nielen, Mol, Sikkema-de Jong, & Bus, 2016; Rubinsten et al., 2015). A measure for the level of MA is obtained by comparing the reaction times on a simple task that is presented right after a MA evoking or neutral stimulus. The simple task can either be at the same side or the opposite side of the screen as the MA evoking or neutral stimulus. The MA evoking stimulus influences the reaction time for the simple task in two ways (Rubinsten et al., 2015). First, it negatively influences the reaction time in trials where the task is on the opposite side of the screen as the MA evoking stimulus due to slow attention disengagement. Secondly, it positively influences the reaction time in trials where the task is on the same side of the screen as the MA evoking stimulus due to attentional engagement.

Current Study

In light of the gender gap in enrolled students in STEM studies and the proposed importance of MA in the pursuit of a STEM major and career, MA and its related variables should be investigated (Ashcraft & Ridley, 2005). The current study aims to investigate the difference in MA between male and female students in their first year of the preacademic educational track (HAVO, atheneum and gymnasium) in secondary school in the Netherlands. To investigate this, multiple methods will be used to assess MA to rule out possible bias from self-report questionnaires. Due to the COVID-19 pandemic, the study had to be revised into an exploratory study with fewer participants whom cannot be assumed to be representative of the population of first year students of the preacademic educational track (HAVO, atheneum and gymnasium) in secondary schools in the Netherlands. The results in the current study cannot be generalized to the population.

The research question concerns the gender differences and different assessment methods for MA: "*Is there a difference in the population of male and female students with high MA, when MA is measured by self-report questionnaires or using a Numerical Dot-Probe Task, controlling for working memory, mathematical achievement and general anxiety?*". First, it is hypothesized that there will be more girls with high levels of MA compared to boys when looking at the self-report questionnaires. However, the self-report questionnaires are expected to be susceptible to biases that might cause an overrepresentation of girls and an underrepresentation of boys. This leads to the second hypothesis that there will be no difference between the number of boys and girls with high levels of MA when looking at the Numerical Dot-Probe Task. Concepts that are possibly related to MA will be controlled for. Based on prior studies, it is hypothesized that MA and mathematical achievement are negatively related, a student with higher levels of MA will have lower mathematical achievement (Hembree, 1990; Suárez-Pellicioni et al., 2016). Both visual and verbal working memory are expected to be equally negatively related to MA (Miller & Bichsel, 2004). It is expected that general anxiety will have a positive relationship with MA (Hembree, 1990). However, when comparing general anxiety and MA, the variance of one of the two constructs cannot completely be predicted from the variance of the other indicating that MA is different from general anxiety.

Methods

Participants

The participants for the current study were recruited using convenience sampling. A high school for preacademic secondary education (HAVO, atheneum and gymnasium), located in the Randstad in the Netherlands, agreed to cooperate. Informed consent and information regarding the study were distributed to the first-year students. Since there were no exclusion criteria, all first-year students could participate. The final sample consisted of seven students (two girls and five boys). These seven participants cannot be assumed to be representative of the population of first year secondary school students in the higher, preacademic educational track (HAVO, atheneum and gymnasium). In the sample, one student received education on HAVO level, three students on atheneum level and three students on a mixed level. On the informed consent, students indicated the presence or absence of previously classified dyslexia or dyscalculia ($n_{\text{dyslexia}} = 2$; $n_{\text{dyscalculia}} = 0$).

Procedure

Before the start of data collection, the informed consent had to be signed by the parents or guardians with consent of the student that agrees to participate. The data collection consisted of two sessions of approximately 30-40 minutes at their school. For the first session, small groups of students were taken out of their classrooms to individually complete the Numerical Dot-Probe Task and three short questionnaires to assess MA and general anxiety. In the following week, the second session to assess the students' working memory was planned. The working memory assessment was administered to the whole group at once but the researchers and teachers checked that the tasks were made individually. In another week, during school hours, the mathematical performance of the students was assessed by their mathematics teacher. After the data collection, the mathematics teachers received a small thank-you present to enhance their instruction. The study was approved by the ethics committee of the Institute Education and Child studies of Leiden University (ECPW2019-246).

Data-analysis

In order to answer the research question, "*Is there a difference in the population of male and female students with high MA, when MA is measured by self-report questionnaires or using a Numerical Dot-Probe Task, controlling for working memory, mathematical achievement and general anxiety?*", the z-scores of all participants on the different tests for all variables were compared to examine whether they prove or disprove the hypotheses. The adapted t-test was used to identify scores that differed significantly from the other scores for each measurement (Crawford & Garthwaite, 2002; Crawford, Garthwaite, & Wood, 2010; Crawford & Howell, 1998).

Measures

Mathematical performance. Mathematical performance was acquired using two mathematics tests of 19 questions administered at school by the teacher. The test was developed by the school and based on the Dutch curriculum 'Getal & Ruimte' [Number & Space] and is a standard test in the curriculum of the students. The tests consisted of basic calculations with and without context, fractions, converting measurements and times. Students were not allowed to use their calculator, the test thus assessed their actual mathematical skills and reasoning skills. Teachers provided the scores for each participant to the researchers.

The Numerical Dot-Probe Task (NDPT). The Numerical Dot-Probe Task was a computer-based test that used the concept of attentional bias in MA (Bar-Haim et al., 2007). The development of the Numerical Dot-Probe Task for the current study was based on an earlier study by Rubinsten et al. (2015). The Numerical Dot-Probe Task consisted of four blocks of 12 trials (64 trials total), two blocks were separated by a short, relaxing, one-minute video of an aquarium. All trials had the same outline but could be either mathematics related or neutral, see Figure 1 for a schematic overview of a mathematics related trial.

For each trial, the attention of the participant was drawn to a neutral stimulus (a neutral word such as "Citroen" [Lemon]) or mathematics related stimulus (either an equation such as "46/23" or a mathematics related word such as "Optellen" [Counting]). For a complete list of the stimuli, see Appendix A. This stimulus, or prime, appeared at one side of the computer screen and was followed by a probe (either one or two asterisks "*") that was presented at either the same side of the screen (congruent) or the opposite side of the screen (incongruent). For the first task, the participant quickly pressed a key indicating the number of asterisks shown, 'z' for one asterisk and 'm' for two asterisks. After the probe, there was a second task; the participant was shown either a number which could be the answer to the earlier shown equation (mathematics related item) or a new word that could rhyme with the previously shown word (mathematics related item or neutral item). The participant needed to indicate whether the number is the correct or incorrect answer or whether the word does or does not rhyme by using the same keys, 'z' for a correct answer and 'm' for an incorrect answer. By comparing the reaction times to the probe after a mathematics related prime and a neutral prime, a measure for the level of MA is obtained (Rubinsten et al., 2015). Higher levels of MA in the individual were assumed when two requirements were met; (1) faster responses on the noncongruent, mathematics related trials, and (2) no big difference between congruent and noncongruent trials on the non-mathematics related probes. The Numerical Dot-Probe Task takes 25 minutes and is thought to be a valid and reliable measurement for the level of MA (Rubinsten et al., 2015).

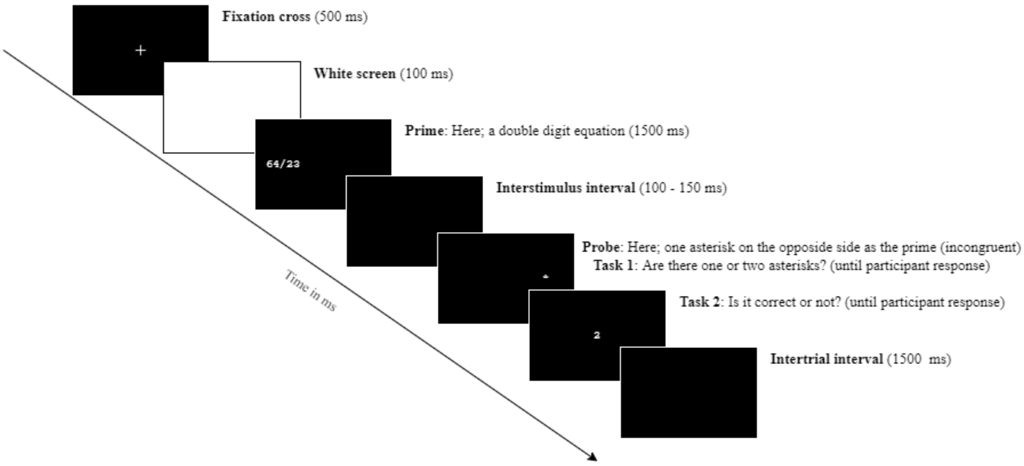

Figure 1. Example of a mathematics related trial in the Numerical Dot-Probe Task.

Rekenbelevingsschaal (RBS). The Rekenbelevingsschaal *[Mathematics Experience Scale]* was a self-report questionnaire designed to measure the level of MA in an individual (van der Beek, Toll, & van Luit, 2017). On a four point Likert scale, participants first indicated how often he/she thinks, feels or does something when having difficulty solving an equation (van der Beek et al., 2017). For example, the participant could indicate whether he/she never, sometimes, often or always tries to solve the easier equation first in this scenario. Secondly, participants used a four point Likert scale to indicate how much he/she agreed with the statement (van der Beek et al., 2017). For example, the participant could completely disagree, disagree, agree or completely agree with the statement that he/she feels anxious when solving equations. Higher scores on the Mathematics Experience Scale indicated higher levels of MA in the individual. The Mathematics Experience Scale contains 66 items, took 5-10 minutes and gave a valid and reliable indication of MA for students in ‘groep’ 6 through 8 of primary education and ‘klas’ 1 through 3 of secondary education (COTAN documentatie, 2017; van der Beek et al., 2017). The Mathematics Experience Scale was originally a paper and pencil questionnaire, but was administered on the computer for this study. For the current study and its sample, the Crohnbach’s Alpha for the Mathematics Experience Scale was $\alpha = .9$.

The translated Modified Abbreviated Math Anxiety Scale (mAMAS-NL). The mAMAS-NL was a self-report questionnaire designed to measure the level of MA in an individual (Carey, Hill, Devine, & Szücs, 2017; Hopko, Mahadevan, Bare, & Hunt, 2003). For the purpose of this study, the mAMAS was translated into Dutch (i.e. mAMAS-NL). Researchers used the knowledge of a native English professor at Leiden University and the native Dutch mathematics teacher to create the best translation for the current setting. The questionnaire consisted of nine statements relating to experiences within the context of mathematics and took a maximum of 5 minutes. Participants indicated on a 5-point Likert scale how anxious he/she would feel in the described situation (Carey et al., 2017). Higher scores on the mAMAS-NL indicated higher levels of MA in the individual. The mAMAS provided a valid and reliable scale for MA in children and adolescents (Carey et al., 2017). The mAMAS-NL was administered on the computer for this study. For the current study and its sample, the Crohnbach’s Alpha for the translated Modified Abbreviated Math Anxiety Scale was $\alpha = .68$.

Revised Children Anxiety and Depression Scale (RCADS47-NL). The Revised Children Anxiety and Depression Scale was a self-report questionnaire designed to measure general anxiety in an individual (Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000). On a four-point Likert scale, participants indicated how applicable each of the 47 statements were to them. Higher scores indicated higher levels of general anxiety in the individual. The statements touch on different types of worrying thoughts students might have. The assessment took 5-10 minutes and was a valid and reliable measure of general anxiety in (Dutch) children and adolescents from different ethnical backgrounds (Chorpita et al., 2000; Kösters, Chinapaw, Zwaanswijk, van der Wal, & Koot, 2015; Muris, Meesters, & Schouten, 2002; Piqueras, Martín-Vivar, Sandin, San Luis, & Pineda, 2017). The Revised Children Anxiety and Depression Scale was originally a paper and pencil questionnaire, but was administered on the computer for this study. For the current study and its sample, the Crohnbach's Alpha for the Revised Children Anxiety and Depression Scale was $\alpha = .96$.

Working Memory. Two working memory tests of the Wechsler Intelligence Scale for Children were adapted and used to assess the working memory of the participants (Wechsler, 2018). The *working memory number task* (WM-digits) consisted of two parts with 9 items per part. For both parts, participants heard two to nine numbers depending on the difficulty. For the first part, the reverse number task, participants had to write the numbers down in reverse order. For the second part, the sequencing number task, the participants wrote the numbers down from low to high value. The *working memory picture task* (WM-pictures) was the second test which consisted of 19 items. For this test, the participants had five seconds to view and memorize two to five pictures in the correct order, depending on the difficulty. On the next screen, the participants were shown multiple pictures with a letter beneath each picture. Some of the pictures were not part of the earlier shown set. Students wrote down the letters of the earlier seen pictures in the correct order. The number of pictures on both screens increased with every series of pictures. In total, the assessment of working memory took 20-30 minutes. Students received two points for a set in the correct order and one point for a set containing all correct numbers or letters but in the wrong order, a higher score indicated a higher working memory capacity of the individual. The working memory number and picture tasks are reliable measures for working memory in children and adolescents (Wechsler, 2018).

Results

Data exploration

All seven participants completed the questionnaires for MA and general anxiety (see Table 1). One participant was absent for the assessment of the working memory tasks in the second session. This participant will not be included in the analyses that include working memory. For the score of mathematical performance, researchers acquired the grades for two different tests for all participants (*N* = 7). Depending on the level of education of the student, each test was graded by their own teacher using either or both a HAVO norm (*n* = 4) and/or atheneum norm (*n* = 6). Using the grades of the participants with both HAVO and atheneum grades (*n* = 3), the mean differences between the HAVO and atheneum grades were calculated for each test. For each test, HAVO grades were on average 1.47 (test 1) and 1.2 (test 2) points higher than the atheneum grade. The mean differences and HAVO grades were used to estimate the atheneum grades for both tests. A mean grade for mathematical performance was calculated for all participants based on the atheneum normed grades (see Table 1).

Table 1
Descriptive statistics for questionnaires, working memory and mathematical performance

Test	*n*	*M(SD)*
Mathematical performance	7	6.93(1.49)
mAMAS-NL	7	19.14(4.85)
RBS	7	14.86(7.84)
RCADS47-NL	7	24.57(17.15)
WM-digits	6	9.33(2.34)
WM-pictures	6	26.33(16.67)

Note. mAMAS-NL = modified Abbreviated Math Anxiety Scale in Dutch; RBS = Mathematics Experience Scale; RCADS47-NL = Revised Children Anxiety and Depression Scale in Dutch; WM-digits and pictures = Working Memory for either digits or pictures.

In line with the mAMAS-NL manual, the sum of all items resulted in a MA score for all seven participants (*M* = 19.14; *SD* = 4.85; see Table 1). For the RBS, the manual was consulted to identify the 14 items that should be summed for a score for MA (*M* = 14.86;

SD = 7.84; see Table 1). The manual of the RCADS47-NL was used to identify the 37 items that had to be summed in order to calculate an anxiety score for all seven participants (*M* = 24.57; *SD* = 17.15; see Table 1). For WM-digits and WM-pictures, the scores on all items were summed for their final score ($M_{WM\text{-}digits}$ = 9.33; $SD_{WM\text{-}digits}$ = 2.34; $M_{WM\text{-}pictures}$ = 26.33; $SD_{WM\text{-}pictures}$ = 16.67; see Table 1).

Numerical Dot-Probe Task (NDPT)

Overall, participants answered 87.7% of primes correctly, ranging from 82.8% to 92.2% for each participant. In line with the literature, incorrect trials (12.3%) were removed from further analyses (Rubinsten et al., 2015). The ten prime categories were: single equation addition, single equation subtraction, single equation multiplication, single equation division, double equation addition, double equation subtraction, double equation multiplication, double equation division, mathematics related words and non-mathematics related words. In most prime categories there was a score of 80% correctly answered primes, single equation subtraction and single equation division had the highest percentages with 96.4%. The prime categories with the lowest percentage correct that were still above 80% were single equation multiplication and double equation addition with 82.1%. Only the category double equation multiplications had less than 80% correctly answered primes (46.4%), this category will be removed from further analyses (Rubinsten et al., 2015).

Next, reaction times (milliseconds; ms) to the probes following the primes were examined. The percentage of outliers for each participant ranged from 0% to 8.47%. Outliers were winsorized (Sullivan, Warkentin, & Wallace, 2021). The remaining, winsorized, reaction times (*n* = 418) were inspected (see Table 2). The mean reaction time for mathematics related trials ranged from 1221.06 to 2491.78 milliseconds, depending on the type of mathematics related prime. Due to the large differences in means and standard deviations of these reaction times to probes in the different prime categories, mathematics related trials were grouped to explore different ways of estimating attentional bias. The five mathematics related groups were: all mathematics related probes, mathematics related words, all mathematics equations, single digit equations and double digit equations (see Table 3).

Table 2
Reaction times (in milliseconds) to probe categories

Probe category	*n*	*M(SD)*
Single equation addition	28	1316.86 (721.83)
Single equation subtraction	27	1267.15 (939.89)
Single equation multiplication	28	1453.25 (887.88)
Single equation division	28	1352.93 (936.77)
Double equation addition	27	2491.78 (1726.38)
Double equation subtraction	28	2307.46 (1769.71)
Double equation division	28	1790.96 (952.92)
Mathematics related words	112	1221.06 (812.29)
Non mathematics related words	112	1147.96 (924.23)

Table 3
Reaction times (in milliseconds) to probe category groups

Probe category	*n*	*M(SD)*
All mathematics	306	1530.89 (1145.91)
Mathematics related words	112	1221.06 (812.29)
Mathematics equations	194	1709.75 (1267.98)
Single digit equations	111	1348.27 (866.18)
Double digit equations	83	2193.18 (1537.62)
Non mathematics related words	112	1147.96 (924.23)

The effect sizes (Cohen's *d*) of six independent t-tests between the reaction times of congruent and incongruent trials for all mathematics related and neutral probe groups were used to estimate attentional bias. A participant is assumed to show attentional bias if they meet two requirements; (1) a positive effect size on the mathematics related probe to indicate faster responses on the noncongruent trials, and (2) a lower effect size on the non-mathematics related probes to indicate no big difference between congruent and noncongruent trials. Two probe category groups (all mathematics and mathematics equations) were deleted since the other probe category groups already include all trials. Participants received four binominal (0 or 1) scores for attentional bias, one per mathematics related probe category group (see Table 4). A score of '1' indicates attentional bias and a score of '0' indicates no attentional bias. The final score for attentional bias in the NDPT for a participant is the sum of the probe category scores.

Table 4
Indications for attentional bias for each participant and probe category group

Participant	Mathematics related words	Single digit equations	Double digit equations	Total score
BOY_1	1	0	1	2
GIRL_1	0	0	0	0
BOY_2	0	0	0	0
GIRL_2	0	0	0	0
BOY_3	0	0	0	0
BOY_4	1	0	0	1
BOY_5	0	1	0	1

Table 5
Total scores (and accompanying z-scores in brackets) for each measurement per participant.

Partici-pant	mAMAS-NL	RBS	NDPT	RCADS 47-NL	Math Achieve-ment	WM_ digits	WM_ pictures
GIRL_1	16(-.65)	12(-.36)	0(-.73)	15(-.56)	6.22(-.48)	12(1.14)	23(-.82)
GIRL_2	24(1.00)	22(.91)	0(-.73)	32(.43)	6.55(-.26)	10(.29)	27(.16)
BOY_1	22(.59)	26(1.42)	2*(1.82)	54(1.72)	7.4(.32)	7(-1.00)	23(-.82)
BOY_2	13(-1.27)	6(-1.13)	0(-.73)	32(.43)	8.7(1.19)	12(1.14)	34*(1.88)
BOY_3	25(1.21)	18(.40)	0(-.73)	11(-.79)	4.5(-1.63)	8(-.57)	26(-.08)
BOY_4	14(-1.06)	5(-1.26)	1(.54)	2(-1.32)	6.45(-.32)	7(-1.00)	25(-.33)
BOY_5	20(.18)	15(.02)	1(.54)	26(.08)	8.7(1.19)	-	-

Note. mAMAS-NL = modified Abbreviated Math Anxiety Scale in Dutch; RBS = Mathematics Experience Scale; RCADS47-NL = Revised Children Anxiety and Depression Scale in Dutch; WM-digits and pictures = Working Memory for either digits or pictures.
* $p < .05$.

Research question

The research question in the current study was: "*Is there a difference in the population of male and female students with high MA, when MA is measured by self-report questionnaires or using a Numerical Dot-Probe Task, controlling for working memory, mathematical achievement and general anxiety?*". Multiple hypotheses were stated, these hypotheses will be discussed using the total scores and *z*-scores of all measures shown in Table 5. The adapted t-test was used to identify scores that differed significantly from the other scores for each measurement (Crawford & Garthwaite, 2002; Crawford et al., 2010; Crawford & Howell, 1998).

First, it was hypothesized that there will be more girls with high levels of MA compared to boys when looking at the self-report questionnaires. When looking at the final scores and the *z*-scores of the self-report questionnaires (mAMAS-NL and RBS), girl GIRL_2 reported higher levels of MA compared to the average of the sample ($z_{\text{mAMAS-NL}} = 1.00$; $z_{\text{RBS}} = .91$). Overall, one of two girls (50%) and three of five boys (60%) show high levels of MA when looking at the self-report questionnaire, this does not confirm the hypothesis. The self-report questionnaires were expected to be susceptible to biases that might have caused an overestimation of girls' MA score and an underestimation of boys' MA score. Both female participants, GIRL_2 ($z_{\text{mAMAS-NL}} = 1.00$; $z_{\text{RBS}} = .91$; $z_{\text{NDPT}} = -.73$) and GIRL_1 ($z_{\text{mAMAS-NL}} = -.65$; $z_{\text{RBS}} = -.36$; $z_{\text{NDPT}} = -.73$), reported higher levels of MA when looking at the self-report questionnaires compared to their score on the NDPT. There was a male participant, BOY_3, who showed the same pattern as the girls, higher levels of MA when looking at the self-report questionnaires compared to his score on the NDPT ($z_{\text{mAMAS-NL}} = 1.21$; $z_{\text{RBS}} = .40$; $z_{\text{NDPT}} = -.73$). Overall, two of two girls (100%) and only one of five boys (20%) showed higher levels of MA when looking at their scores on the self-report questionnaires compared to their score on the NDPT. These observations confirm the hypothesis.

The second hypothesis stated that there is no difference between the number of boys and girls with high levels of MA when looking at the NDPT. Both girls have a *z*-score of -.73 for the NDPT and thus 0% of the girls have high levels of MA when looking at the NDPT. One male participant (BOY_1) scored significantly higher compared to the mean of the sample, this is a boy ($z_{\text{NDPT}} = 1.82$). This shows that one out of five, or 20%, of boys have high levels of MA when looking at the NDPT. These observations contradict the hypothesis.

When looking at the control variables, it was hypothesized that MA and mathematical achievement were negatively related, a student with higher levels of MA would have lower mathematical achievement (Hembree, 1990; Suárez-Pellicioni et al., 2016). On top of this, both visual and verbal working memory were expected to be equally negatively related to MA

(Miller & Bichsel, 2004). These hypotheses could not be confirmed or denied. Participant BOY_2 confirms both of these hypotheses, he showed low scores for MA for both self-report questionnaires and the NDPT ($z_{\text{mAMAS-NL}}$ = -1.27; z_{RBS} = -1.13; z_{NDPT} = -.73) and high scores for WM and mathematical achievement ($z_{\text{WM\_digits}}$ = 1.14; $z_{\text{WM\_pictures}}$ = 1.88; $z_{\text{MathAchievement}}$ = 1.19). Participant BOY_3 also confirmed both hypotheses, he showed slightly higher scores for MA on both the self-report questionnaires and the NDPT ($z_{\text{mAMAS-NL}}$ = 1.21; z_{RBS} = .40; z_{NDPT} = -.73) and lower scores on WM and mathematical achievement ($z_{\text{WM\_digits}}$ = -.57; $z_{\text{WM\_pictures}}$ = -.08; $z_{\text{MathAchievement}}$ = -1.63). Participant BOY_1 partly confirms the hypotheses; he has high scores on MA on both the self-report questionnaires and the NDPT ($z_{\text{mAMAS-NL}}$ = .59; z_{RBS} = 1.42; z_{NDPT} = 1.82) and low scores on WM ($z_{\text{WM\_digits}}$ = -1.00; $z_{\text{WM\_pictures}}$ = -.82), however he has a slightly higher score on mathematical achievement ($z_{\text{MathAchievement}}$ = .32) which is not in line with the hypothesis. Overall, most participants did not definitively show a strong negative relationship between MA and mathematical achievement and both types of WM. However, the scores did indicate a possible pattern of a negative relationship between MA and mathematical achievement and WM.

Finally, it was expected that general anxiety would have a positive relationship with MA (Hembree, 1990). This hypothesis was partly confirmed. The scores of general anxiety (RCADS47-NL) and MA (mAMAS-NL, RBS and NDPT) for three participants (BOY_1, GIRL_1 and BOY_5) showed that when either general anxiety or MA is elevated, the final score on the other construct was elevated as well. For example, participant BOY_1 scored above the average of the current sample for general anxiety ($z_{\text{RCADS47-NL}}$ = 1.72) and MA ($z_{\text{mAMAS-NL}}$ = .59; z_{RBS} = 1.42; z_{NDPT} = 1.82).

Further observations

Due to the exploratory nature of the study, some other trends in the data will be discussed. First, the *z*-scores of all participants for the two different MA self-report questionnaires (RBS and mAMAS-NL) are in the same range. This means that a participant with a high score on the RBS also has a high score on the mAMAS-NL. The self-report questionnaires for MA are related and likely to measure the same or a similar construct.

Second, the two different indicators for WM were in the same range for all participants except for participant GIRL_1. Participant GIRL_1 showed a large difference between the two indicators relative to the other participants of the sample ($z_{WM\_digits} = 1.14$; $z_{WM\_pictures} = -.82$). Compared to the other participants in the sample, participant GIRL_1 scored the highest on WM_digits and the lowest on WM_pictures. Participant GIRL_1 also had lower scores for MA, general anxiety and mathematics achievement compared to the other participants.

Discussion

Considering the gender gap in students enrolled in STEM studies and the possible importance of MA in the pursuit of a STEM major and career, MA and its related variables were investigated in the current study (Ashcraft & Ridley, 2005). Multiple methods were used to assess MA to limit the possible bias from self-report questionnaires. The research question was formulated as follows: "*Is there a difference in the population of male and female students with high MA, when MA is measured by self-report questionnaires or using a Numerical Dot-Probe Task, controlling for working memory, mathematical achievement and general anxiety?*". In order to answer this question, multiple hypotheses were investigated. Due to the COVID-19 pandemic, the original study had to be revised into an exploratory study with seven participants whom cannot be assumed to be representative of the population of first year secondary school students in the higher, preacademic educational track (HAVO, atheneum and gymnasium). We therefore discuss the results in a more qualitative way.

First, the hypothesis that there will be more girls with high levels of MA compared to boys when looking at the self-report questionnaires was partly confirmed. In the study we found that less girls showed high levels of MA compared to boys when looking at the self-report questionnaires. However, the self-report questionnaires were susceptible to biases since the girls' MA score was overestimated and the boys' MA score was underestimated when comparing the self-report questionnaires to a MA measure that is considered less prone to social desirability (van de Mortel, 2008; Rubinsten en al., 2015). This effect was not strong enough to result in an overrepresentation of girls and an underrepresentation of boys in the population of children with high levels of MA. While one boy has high levels of MA when looking at the implicit MA task, the Numerical Dot-Probe Task, neither of the girls have high levels of MA.

When looking at the control variables, it is hypothesized that MA and both mathematical achievement and WM are negatively related. That is, a student with higher levels of MA will have lower mathematical achievement (Hembree, 1990; Suárez-Pellicioni et al., 2016). On top of this, both visual and verbal WM are expected to be equally negatively related to MA (Miller & Bichsel, 2004). The results indeed show this pattern. Finally, it is expected that general anxiety has a positive relationship with MA (Hembree, 1990). This hypothesis was partly confirmed since three participants showed a positive relationship between general anxiety (RCADS47-NL) and MA (mAMAS-NL, RBS and NDPT). Overall, the control variables showed the expected patterns.

The main aim of the current study was to investigate the difference in MA between male and female students in their first year of the preacademic educational track (HAVO, atheneum and gymnasium) in secondary school in the Netherlands using different measurement methods. The first two hypotheses concerning gender differences were not fully confirmed in the current study, this unexpected result cannot be explained by control variables since they showed the expected patterns. Further explanations will be explored.

Prior research has shown that female students without MA are likely to report MA in self-report questionnaires due to child- (e.g., gender identification and experienced mathematical abilities) and societal (e.g., stereotype threat) factors (Ashcraft, 2002; van de Mortel, 2008). An important factor that influences individual variability in reporting MA is stereotype threat (Maloney, Schaeffer, & Beilock, 2013; Schmader, Johns, & Barquissau, 2004). Stereotype threat can influence MA in an individual, but especially women who value mathematics, identify strongly with their female identity and are aware of the stereotype that women perform lower on mathematics experience negative consequences of the stereotype threat (Schmader et al., 2004). It is possible that neither of the female participants have high levels of MA but reported MA in the self-report questionnaires due to these child and societal factors.

Another way stereotype threat influences the gender gap in STEM studies is through the male stereotyped views (Hyde, Fennema, Ryan, Frost, & Hopp, 1990). The meta-analysis by Hyde et al. (1990) showed that more males than females hold stereotyped views such as 'men are better at mathematics' and 'females who achieve in mathematics are less feminine'. The males' stereotyped views can be problematic when it encourages females to not achieve in mathematics by subtly hinting to females that females achieving in mathematics are less feminine. Males' stereotyped views might lead to male teachers actively discouraging girls from taking STEM courses or male employers turning female applicants away from STEM related jobs. These hypotheses should be investigated in future studies.

The current study has not included a variable indicating the values of the students regarding mathematics, their gender identity or their awareness of the stereotype that women perform lower on mathematics. The effect of these factors can thus not be investigated but could impact the MA scores significantly since men and women that are more aware of these stereotypes and identify more with the groups involved in the stereotype are more affected by the stereotype (Schmader et al., 2004). Even though the current study has not investigated stereotype threat, it is possible that interventions that target working memory can improve both MA and the possible stereotype threat since it is thought that MA and stereotype threat effect mathematical performance through similar mechanisms, namely working memory capacity (Maloney et al., 2013). The absence of female students with high levels of MA in the current sample and the small total sample size could explain why the effect of self-report bias was not strong enough to result in an overrepresentation of girls and an underrepresentation of boys in the population of students with MA.

The current study suggests that the prominent gender gap in STEM studies might not be due to gender differences in MA since the gender differences are smaller when using less biased measures. Due to the small sample size of the current study, it is important that these findings are replicated in future studies using bigger, more representative, samples. Stereotype threat is a variable that could explain the STEM gender gap and should thus be included in these future studies.

References

Aguinis, H., Ji, Y. H., & Joo, H. (2018). Gender productivity gap among star performers in STEM and other scientific fields. *Journal of Applied Psychology, 103*(12), 1283-1306. doi:10.103/apl0000331

Ahmed, W. (2018). Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. *Journal of Adolescence, 67*, 158-166. doi:10.1016/j.adolescence.2018.06.010

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences, 22*(3), 385-389. doi:10.1016/j.lindif.2011.12.004

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th. ed.). Arlington, VA: Author.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. doi:10.1111/1467-8721.00196

Ashcraft, M. H., & Faust, M W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition and Emotion, 8*(2), 97-125. doi:10.1080/02699939408408931

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*(2), 224-237. doi:10.1037//0096-3445.130.2.224

Ashcraft, M. H., Kirk, E. P., Hopko, D. (1998). On the cognitive consequences of mathematics anxiety. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 175-196). Hove, England: Psychology Press Ltd., Publishers

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27*(3), 197-205. doi:10.1177/0734282908330580

Ashcraft, M. H., & Ridley, K. (2005). Math anxiety and its cognitive consequences: A tutorial review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 315-325). New York, NY: Psychology Press

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1-24. doi:10.1037/0033-2909.133.1.1

van der Beek, J. P. J., Toll, S. W. M., & Van Luit, J. E. H. (2017). *RBS Rekenbelevingsschaal* [Mathematics Experience Scale]. Amsterdam: Hogrefe

Beilock, S. L., & Maloney, E. A. (2015). Math anxiety: A factor in math achievement not to be ignored. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 4-12. doi:10.1177/2372732215601438

Benbow, C. P. (2012). Identifying and nurturing future innovators in science, technology, engineering, and mathematics: A review of findings from the study of mathematically precocious youth. *Peabody Journal of Education, 87*(1), 16-25. doi:10.1080/0161956X.2012.642236

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles, 72*(11-12), 536-546. doi:10.1007/s11199-015-0486-9

Brown, C., & Corcoran, M. (1997). Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics, 15*(3), 431-465. doi:10.1086/209867

Black, D. A., Haviland, A., Sanders, S. G., & Taylor, L. J. (2008). Gender wage disparities among the highly educated. *Journal of Human Resources, 43*(3), 630-659. doi:10.1353/jhr.2008.0030

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature, 55*(3), 789-865. doi:10.1257/jel.20160995

Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology, 8*(11), 1-13. doi:10.3389/fpsyg.2017.00011

Centraal Bureau voor de Statistiek [Central Bureau of Statistics]. (March 20, 2020). *Hoger onderwijs; eerste- en ouderejaarsstudenten, studierichting* [Higher education; first-year and senior students, studyfield] [Data portal]. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83538NED/table?ts=1582627214932

Chorpita, B. F., Yim, L. M., Moffitt, C. E., Umemoto L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy, 38*(8), 835-855. doi:10.1016/S0005-7967(99)00130-8

COTAN documentatie. (2017). *Rekenbelevingsschaal, RBS*. [Mathematics Experience Scale] Retrieved from https://www-cotandocumentatie-nl.ezproxy.leidenuniv.nl:2443/beoordelingen/b/15643/rekenbelevingsschaal/

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40*(8), 1196-1208. doi:10.1016/S0028-3932(01)00224-X

Crawford, J. R., Garthwaite, P. H., & Wood, L. T. (2010). Inferential methods for comparing two single cases. *Cognitive Neuropsychology, 27*(5), 377-400. doi:10.1080/02643294.2011.559158

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12*(4), 482-486. doi:10.1076/clin.12.4.482.7241

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, *82*(3), 766-779. doi:10.1111/j.1467-8624.2010.01529.x

Egloff, B., & Schmukle, S. C. (2004). Gender differences in implicit and explicit anxiety measures. *Personality and Individual Differences, 36*(8), 1807-1815. doi:10.1016/j.paid.2003.07.002

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance, The processing efficiency theory. *Cognition and Emotion, 6*(6), 409-434. doi:10.1080/02699939208409696

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46. doi:10.2307/749455

Hopko, D. R., Ashcraft, M. H., Gute, J., Ruggiero, K. J., & Lewis, C. (1998). Mathematics anxiety and working memory: Support for the existence of a deficient inhibition mechanism. *Journal of Anxiety Disorders*, *12*(4), 343-355. doi:10.1016/S0887-6185(98)00019-X

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178–182. doi:10.1177/1073191103010002008

Huang, C., Liao, H., & Chang, S. (1998). Social desirability and the clinical self-report inventory: Methodological reconsideration. *Journal of Clinical Psychology, 54*(4), 517-528. doi:10.1002/(SICI)1097-4679(199806)54:4<517::AID-JCLP13>3.0.CO2-I

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*(3), 299-324. doi:10.1111/j.1471-6402.1990.tb00022.x

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing, 17*(2), 79-103. doi:10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0

Kösters, M. P., Chinapaw, M. J. M., Zwaanswijk, M., van der Wal, M. F., & Koot, H. M. (2015). Structure, reliability, and validity of the revised child anxiety and depression scale (RCADS) in a multi-ethnic urban sample of Dutch children. *BMC Psychiatry,* 15, 132-140. doi:10.1186/s12888-015-0509-7

Kurtz-Costes, B., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender stereotypes about mathematics and science and self-perceptions of ability in late childhood and early adolescence. *Merrill-Palmer Quarterly, 54*(3), 386-409. doi:10.1353/mpq.0.0001

Lou, W., Hogan, D., Tan, L. S., Kaur, B., Ng, P. T., & Chan, M. (2014). Self-construal and students' math self-concept, anxiety and achievement: An examination of achievement goals as mediators. *Asian Journal of Social Psychology, 17*(3), 184-195. doi:10.111/ajsp.12058

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540. doi:10.2307/749772

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15-20. doi:10.1037/0021-843X.95.1.15

Maloney, E. A., Schaeffer, M. W., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences and promising interventions. *Research in Mathematics Education, 15*(2), 115-128. doi:10.1080/14794802.2013.797744

McDonald, K. S., & Waite, A. M. (2019). Future directions: Challenges and solutions facing career readiness and development in STEM fields. *Advances in Developing Human Resources, 21*(1), 133-138. doi:10.1177/1523422318814552

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, and math performance. *Personality and Individual Differences, 37*(3), 591-606. doi:10.1016/j.paid.2003.09.029

Miyazaki, Y., Ichihara, S., Wake, H., & Wake, T. (2012). Attentional bias to direct gaze in a dot-probe paradigm. *Perceptual and Motor Skills, 114*(3), 1007-1022. doi:10.2466/21.07.24.PMS.114.3.1007-1022

Moakler, M. W., & Kim, M. M. (2014). College major choice in STEM: Revisiting confidence and demographic factors. *Career Development Quarterly, 62*(2), 128-142. doi:10.1002/j.2161-0045.2014.00075.x

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

Muris, P., Meesters, C., & Schouten, E. (2002) A brief questionnaire of DSM-IV-defined anxiety and depression symptoms among children. *Clinical Psychology & Psychotherapy, 9*(6), 430-442. doi:10.1002/cpp.347

Namkung, J. M., Peng, P., & Lin, Xi. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459-496. doi:10.3102/0034654319843494

Nielen, T. M. J., Mol, S. E., Sikkema-de Jong, M. T., & Bus, A. G. (2016). Attentional bias toward reading in reluctant readers. *Contemporary Educational Psychology, 46*, 263-271. doi:10.1016/j.cedpsych.2015.11.004

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … 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 – PNAS, 106*(26), 10593-10597. doi:10.1073/pnas.0809921106

Pierce, K. A., & Kirkpatrick, D. R. (1992). Do men lie on fear surveys? *Behaviour Research and Therapy, 30*(4), 415-418. doi:10.1016/0005-7967(92)90055-L

Piqueras, J. A., Martín-Vivar, M., Sandin, B., San Luis, C., & Pineda, D. (2017). The revised child anxiety and depression scale: A systematic review and reliability generalization meta-analysis. *Journal of Affective Disorders, 218*, 153-169. doi:10.1016/j.jad.2017.04.022

Rubinsten, O., Eidlin, H., Wohl, H., & Akibli, O. (2015). Attentional bias in math anxiety. *Frontiers in Psychology*, *6*(1539), 1-9. doi:10.3389/fpsyg.2015.01539

Sass, S. M., Heller, W., Stewart, J. L., Silton, R. L., Edgar, J. C., Fisher, J. E., & Miller, G. A. (2010). Time course of attentional bias in anxiety: Emotion and gender specificity. *Psychophysiology, 47*(2), 247-259. doi:10.1111/j.1469-8986.2009.00926.x

Schmader, T., Johns, M., & Barquissau, M. (2004). The cost of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*(11), 835-850. doi:10.1023/B:SERS.0000029101.74557.a0

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2015). Attentional bias in high-anxious individuals: Evidence from an emotional Stroop task. *Frontiers in Psychology*, *6*(1577), 1-10. doi:10.3389/fpsyg.2015.01577

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*(1), 3-22. doi:10.3758/s13415-015-0370-7

Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research, 132*, 530-543. doi: 10.1016/j.jbusres.2021.03.066

Wang, M. T., Eccles, J. S., Kenny, S. (2013). Not lack of ability but more choice individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*(5), 770-775. doi:10.1177/0956797612458937

Wechsler, D. (2018). *Technische Handleiding WISC-V-NL* [Technical Intstructional Manual WISC-V-NL]. Amsterdam, Nederland: Pearson Benelux.

Appendix A

Stimuli in the Numerical Dot-Probe Task

Equations

For the equations, only single and double digits were included. The equations had a correct and incorrect answer. The incorrect answer was picked as an obvious wrong answer which the participant can identify by only roughly calculating or estimating the original equation. See Table A1 for the equations and their correct and incorrect answers.

Table A1

Single and double digit equations and their correct and incorrect answers

Single digits			Double digits		
Equation	Correct	Incorrect	Equation	Correct	Incorrect
6 + 2	8	16	46 + 23	69	35
6 + 3	9	13	52 + 13	65	60
8 + 2	10	14	64 + 32	96	106
8 + 4	12	6	96 + 48	144	114
6 – 2	4	6	46 – 23	23	18
6 – 3	3	1	52 – 13	39	44
8 – 2	6	3	64 – 32	32	24
8 – 4	4	2	96 – 48	48	62
6 / 2	3	5	46 x 23	1058	1245
6 / 3	2	1	52 x 13	676	520
8 / 2	4	6	64 x 32	2048	1570
8 / 4	2	7	96 x 48	4608	5084
6 x 2	12	16	46 / 23	2	1
6 x 3	18	15	52 / 13	4	6
8 x 2	16	10	64 / 32	2	5
8 x 4	32	18	96 / 48	2	3

Words

In order to calculate the familiarity score, the math related score and the rhyme score for each combination of words, a survey was distributed among acquaintances of the researchers ($N = 59$, $M_{age} = 31,83$, $SD_{age} = 14$ and $range_{age} = 13 – 64$). The *familiarity score* was based on the average answer to the question "Geef op een schaal van 1 -9 aan in hoeverre je bekend bent met de woorden" (1 – 'helemaal niet bekend' and 9 – 'heel bekend'). The *math* *related score* was based on the average answer to the question "Geef op een schaal van 1-9 aan in hoeverre je vindt dat het woord met rekenen of wiskunde te maken heeft" (1 – 'heeft helemaal niet te maken met wiskunde of rekenen' and 9 – 'heeft veel te maken met wiskunde of rekenen'). The rhyme and non-rhyme words that were selected had the same number of syllables as the original word. The *rhyme score* is the average answer to the question "Geef tot slot aan of de volgende combinaties wel of niet rijmen" on a 1 through 3 scale (1 – 'rijmen niet', 2 – 'rijmen een beetje' and 3 – 'rijmen wel'). For the neutral words, the *frightening score* was based on the average score to the question "Geef op een schaal van 1-9 aan in hoeverre de volgende woorden een angstig gevoel bij je oproepen" (1 – roept geen angst op and 9 – heel erg beangstigend). See Table A2 and A3 for the math related and neutral words, their rhyme words, non-rhyme words and average scores.

Table A2

Math related words and their rhyme word, average familiarity score, average math related score and average rhyme score

Word	Rhyme word	Non-rhyme word	Familiarity score	Math related score	Rhyme score
Kwadraat	Bestaat	Vulpen	8.37	8.91	2.73
Grafiek	Ziek	Wereld	6.65	8.53	2.54
Breuken	Deuken	Laptop	8.60	8.84	2.96
Tellen	Bellen	Wachten	8.93	8.73	2.98
Delen	Spelen	Datum	8.81	8.58	2.98
Getal	Val	Titel	8.81	8.75	2.58
Keersom	Andersom	Tijdschrift	7.46	8.55	1.92
Procent	Absent	Slager	8.63	8.64	2.35
Rekenen	Tekenen	Inleiding	8.84	8.64	2.94
Wiskundig	Uitbundig	Wasmanden	8.60	8.58	2.54
Opsommen	Geklommen	Aanbieding	8.35	7.04	2.35
Optellen	Aanstellen	Telefoon	8.82	8.89	2.23
Aftrekken	Klimrekken	Prullenbak	8.82	8.51	2.19
Formule	Capsule	Agenda	8.63	8.69	2.58
Frequentie	Intentie	Volkoren	8.65	7.51	2.63
Cilinder	Vlinder	Restaurant	8.44	6.84	2.44

Table A3

Neutral words and their rhyme word, average familiarity score, average frightening score, average math related score and average rhyme score

Word	Rhyme word	Non-rhyme word	Familiarity score	Frighten-ing score	Math related score	Rhyme score
Vlieger	Bedrieger	Zolder	8.74	1.13	2.33	2.77
Citroen	Schoen	Beschuit	8.81	1.17	1.47	2.71
Bureau	Plateau	Dromen	8.82	1.15	2.20	2.69
Gebruik	Struik	Plinten	8.75	1.02	2.96	2.71
Koffer	Doffer	Bloemen	8.88	1.11	1.51	2.94
Sleutel	Keutel	Wolken	8.89	1.23	2.73	2.88
Piepschuim	Verzuim	Cadeau	8.74	1.34	1.31	2.35
Portret	Besmet	Zolder	8.68	1.09	1.25	2.46
Luidspreker	Aansteker	Boekenkast	8.86	1.36	1.73	2.12
Afscheiding	Opleiding	Badkamer	8.32	2.19	2.33	2.40
Karakter	Compacter	Zonnebloem	8.61	1.30	2.60	2.54
Meubilair	Populair	Pakketje	8.72	1.08	1.58	2.46
Pantoffel	Moffel	Planten	8.68	1.21	1.35	2.65
Aankloppen	Afsoppen	Kalender	8.63	1.70	1.25	2.42
Pepermunt	Verdunt	Kipfilet	8.84	1.06	1.47	2.33
Ovenschaal	Gaspedaal	Pannenkoek	8.65	1.08	1.40	2.62