UNIVERSITY OF PADOVA

Department of Developmental Psychology and Socialisation

Master Degree in Psicologia dello Sviluppo e dell'Educazione Study track in Developmental and Educational Psychology

Final dissertation

Mathematical Domains Across Genders: A Meta-Analytic Investigation

Supervisor
Professor: Sara Caviola

Co-supervisor
Doctor: Lorenzo Esposito

Candidate: Sena Nur Kaya
Student ID number: 2042074

Academic Year: 2023/2024

TABLE OF CONTENTS

INTRODUCTION……………………………………………………………………………4

1. GENERAL ASPECTS OF MATHEMATICAL ACQUISITION ……………………...6

1.1 Mathematical learning……………………………………………………………..6

1.2 Differences in mathematical learning……………………………………………...9

1.3 Gender differences in mathematics………………………………………….............11

2. DIFFERENCES IN MATHEMATICAL ACQUISITION ……………………………15

2.1 Math content………………………………………………………………………..15

2.2 Previous studies…………………………………………………………………….18

3. METHODS………………………………………………………………………………27

3.1 Aims and eligibility criteria………………………………………………………...28

3.1.1 Literature search…………………………………………………………….28

3.1.2 Title and abstract screening…………………………………………………30

3.1.3 Full-text analysis…………………………………………………………….....30

3.1.4 Coding……………………………………………………………………….31

3.2 Analytical strategy………………………………………………………………33

4. RESULTS…………………………………………………………………………….....34

4.1 Forest plot…………………………………………………………………………..35

4.2 Funnel plot……………………………………………………………………….....38

4.3 Analytical results………………………………………………………………….......40

4.4 Math content as a moderator……………………………………………………….41

4.4.1 Advanced math…………………………………………………………….41

4.4.2 Arithmetic…………………………………………………………………....41

4.4.3 Basic numeracy……………………………………………………………...42

4.4.4 Broad mathematics…………………………………………………………..43

4.4.5 Geometry……………………………………………………………………43

4.4.6 National test……………………………………………………………....44

4.4.7 Overview of results……………………………………………………….47

5. DISCUSSION ………………………………………………………………………45

5.1 Findings ………………………………………………………………………...46

5.2 Limitations and future recommendations…………………………………………48

5.3 Conclusion……………………………………………………………………...49

REFERENCES…………………………………………………………………………...50

INTRODUCTION

Mathematics is substantial in understanding and interpreting the real world through internalizing it into our daily lives. Therefore, learning mathematics is useful for acquiring basic concepts and skills for various required areas. Since learning and applying mathematics is very important, previous literature aim to further investigate it, so that authorities and policymakers may consider these factors that may contribute to improve the mathematics learning process and develop innovations and new milestones.

Previous studies indicate that mathematical performance changes depending on various factors such as gender, mathematical content, cultural variations, teaching styles, and such. One of the most controversial aspects is gender. There were several studies on mathematical performance suggesting a male/female advantage. Additionally, this gender gap was changing depending on the mathematical content of the assessment. Thus, this study aimed to investigate mathematical performance between females and males in a meta-analytic perspective within emphasis on mathematical content.

The screening and selection procedures were implemented with PRISMA guidelines, and in conclusion 74 studies published in 2010 and 2016 were extracted and coded, encompassing a total of 508,429 females and 527,857 males. Statistical analysis implemented on the mathematical assessments across gender differences. A gender difference appeared solely on computation assessments in favour of males. However, there was no gender difference in broad mathematics, national tests, geometry, basic numeracy skills, and advanced mathematics abilities. Nevertheless, results indicate that a gender difference in certain mathematics assessments might exist.

This meta-analysis study was useful by detecting the gender gap in math with an emphasis on mathematical content. Indeed, there could be other related factors and moderators that might affect mathematical performance among females and males. Additionally, limitations and further recommendations should be taken in consideration for upcoming studies.

CHAPTER 1

General Aspects of Mathematical Acquisition

1. 1 Mathematical Learning

Mathematics is a field of knowledge that is mostly represented by numbers, symbols, quantities, and abstract concepts to comprehend, perceive, and interpret the real world (Vergnaud, 1987). It is an essential phenomenon that helps to improve core knowledge and skills (Pegg, 2003). According to this study, mathematics is helpful to construct an imaginary understanding of the world with symbolized elements and constructs that are not straightforwardly related with the real world. Learning mathematics is substantial for establishing a base to acquire concepts in various fields such as technology, economics, finance, statistics, social sciences and other related areas (Goldin, 2008). It provides beneficial tools for decision-making, problem solving, and gaining science-related knowledge and skills (Pegg, 2003).

Throughout the literature, there are several studies indicates that mathematics is a broad area involving some specific cognitive domains. In relation with this, Passolunghi (2011) published a study that investigates cognitive elements involved during mathematical learning by contrasting students with mathematical learning disability to students without learning disability through standardized mathematics test. They found that students with mathematical learning disability remained behind in working memory capability, inhibitory skills, speed of processing and displayed higher anxiety scores compared to students without mathematical learning ability. As a result, this study shows these cognitive domains are functioning for a higher mathematics performance. Furthermore, Geary (2011) conducted a longitudinal study on cognitive domains that could be related with mathematics learning ability. Researchers found that number comprehension, number counting, and arithmetic skills assessed in the first year of school were associated with mathematical learning ability until fifth grade, after controlling for domain general cognitive abilities, such as working memory, general intelligence, and processing speed.

The importance of early mathematics education was further addressed by this research showing that students who start school later tend to remain behind in counting, numerical concepts, and simple arithmetic relative to their classmates throughout their education (Duncan et al., 2007; cited by Geary, 2011). In a longitudinal study, Moeller et al. (2011) also discussed that starting early to learn mathematics is useful for further achievement. In this study, they measured basic numerical abilities in the first grade and checked whether they are related to the performance on addition task in the third grade. The results showed that the knowledge that has been practiced continues to accumulate over time.

In summary, research highlights the importance of specific cognitive domains and early mathematics education in mathematics performance. However, there might be multiple other factors of heterogeneous nature that may affect math learning. For instance, Fischer et al. (2011) found a relationship between spatial representation of number magnitude and arithmetic performance and developed a training program for that. They proved that students who participated in the spatial-numerical training program displayed higher scores on a mathematical task. In another study, researchers worked on a study that examines cognitive domains with specialized tools that they improved. They indicated that the central executive, phonological processing, and inhibitory processes thought to be related to early mathematical performances (Navarro et al., 2011). To understand and relate to these early mathematical abilities, a standardized early numeracy test was used. At the end of study, it was found that students who display lower results in central executive, phonological processing, and inhibitory process were also associated with lower mathematics scores.

It is substantial to improve these cognitive domains to internalize better mathematics. Thus, education plays an important role in mathematical learning. However, there are some several challenges among mathematical learning. In accordance with, Hanson et al. (2011) showed that economic challenges, language difficulties, education level of the parents are factors affecting mathematics performance. This study found that children struggle with mathematics if English is not the first language in their house environment and/or if their parents' education level is low. In addition, challenges in the neighborhood environment impact children's academic performance including mathematics. On the other hand, this study revealed that children show higher social capabilities when they live in the English-speaking neighborhoods. Thus, they perform better academically, especially in mathematics and engage in the classroom environment (Hanson et al., 2011). Therefore, it seems to be essential to create an equal learning environment for everyone. Furthermore, Alkharusi (2016) stated that children engage with various kind of assessment tool in the school environment, which might affect their academic achievement, including mathematics. In other words, how well children discern the mathematical assessment they engage in is associated with their performance. Thus, it is important for authorities to take these into consider while preparing syllabuses, assessments, and intervention programs (Alkharusi, 2016). In this regard, Adnan et al., (2014) conducted a study to understand students' engagement, teacher behavior, harmony, equity, students' perspective of connection in the classroom and whether there is an association with academic performance including mathematics. The results demonstrated that students' perspective of the learning environment was the strongest variable that is associated with academic performance, with harmony, equity, teacher behavior, and engagement also played significant roles (Adnan et al.,2014). In general, mathematics plays a crucial role in developing problem-solving skills, and achieving high academic performance. To support all students in acquiring the necessary cognitive skills for mathematics, it is essential to create an equal and inclusive learning environment. This will enable them to perform better and reach their full potential.

1.2 Differences in Mathematics Learning

Academic achievement and mathematic learning performance may be impacted across different cultures. Sandoval-Hernandez & Bialowolski (2016) conducted a comparison study on socially advantaged and disadvantaged students in Singapore, South Korea, Hong Kong, Chinese Taipei, and Japan from a 2011 Trends in International Mathematics and Science Study (TIMSS) study. The study found that students' positive attitudes, teachers' beliefs about students, and students' maternal language being compatible with the task language impacted their mathematics performances positively. Diverse results emerged depending on the nation as students showed higher performances in mathematics as socially advantaged students in Singapore because of families' positive attitudes and engagement with mathematics. In Korea, male students displayed better results in mathematics whereas in Taipei, students displayed better results when they were not confronted with bullying. In relation with this, Mann & DiPrete (2016) discussed the gender distribution in STEM in terms of cross-cultural associations. Researchers highlighted that females show lower performance in STEM related areas compared to males and this gender difference changes with cultural aspects. Females performed higher in nations that are more gender equal compared to nations that are less. Furthermore, Uttal (1997) revealed an interesting viewpoint by conducting a study comparing the mathematics performances of students in the United States and Asian countries such as China, Japan, Taiwan, and others. The recent literature showed a similar result as children in Asia generally perform higher in mathematics than children in the United States. This study aimed to reveal these differences and found that Asians endorsed the idea that being devoted and making attempts are important when learning mathematics, whereas Americans tend to think success comes with innate skills. In turn, this affected students' mathematics performances, in addition to parents' reaction to their children's accomplishments and failures. That is, Americans tend to not pressure students even if they underachieve whereas this situation does not exist for Asian countries (Uttal, 1997). Researchers discussed that the American way of thinking might lead to the viewpoint that they should not work on it if this is an innate ability. Therefore, it is evident that culture may influence learning not due to genetic differences across nations, but rather due to various cultural perspectives and styles.

Mathematical learning and performance may be impacted by cultural variations, showing that mathematics performance can differ for some ethnicities (Boaler et al., 2011). This study extracted some ethnic groups from UK such as African, Bangladeshi, Caribbean, Chinese, Indian, Pakistani, White British. Among them, Chinese and Indian students achieved higher scores in mathematics performance especially compared to Caribbean and Pakistani students. Researchers discussed that educational environment may impact these differences in mathematics, and further attention should be given on this. In accordance with this, Konstantopoulos & Chung (2011) revealed that teacher behaviors might affect students' mathematics performances, especially in schools with minority students. Additionally, immigration was found to be related with the math performance if students are second-generation immigrants in Denmark (Nielsen & Rangvid, 2012). Specifically, students demonstrated lower scores in mathematics depending on their families' residence in the country where they migrated. It was found that students displayed lower scores in mathematics if their fathers spent less year in the country (Nielsen & Rangvid, 2012). Researchers highlighted the importance of challenges of migration on education.

Mathematics learning can be affected by disabilities as well, meaning children with disabilities might be more disadvantaged compared to children without disabilities (Lambert & Tan, 2016). Researchers emphasized teachers' behaviors rather than the health and medical viewpoint.

They found that there were significant differences between students with learning disability and those without learning disability in favour of students without learning disability. Additionally, they underlined the idea that mathematics education should pay attention to these diversities between students with and without learning disabilities. In another study, it was found that children with dyslexia have some challenges in mathematics compared to typically developed children (Cruz-Rodrigues et al., 2014). Similarly, Ross & Randolph (2014) found that children with attention deficit and hyperactivity disorder (ADHD) demonstrated lower scores on a math accuracy test compared to children without ADHD due to their tendency to become easily distracted. Another study on autism spectrum disorder (ASD), showed that students with ASD have difficulties when performing on math-applied tests compared to students without ASD (Oswald et al., 2016). On the other hand, Elbaum (2007) said that low performance in mathematics might be related to language difficulties rather than mathematical abilities for students who have learning disability. In this study, researchers administered oral test for students that have learning disability in mathematics. They found that there were no notable differences among students with and without learning disability in mathematics performances. While some studies support the view that some learning disabilities affect directly mathematics performance, the evidence is not conclusive. Performance differences may be influenced by the teaching environment and learning styles, highlighting the importance of creating inclusive and supportive educational settings.

1.3 Gender differences in Mathematics

Throughout the literature, differences in math performance between females and males as well as a gender gap in math and science related subjects (called STEM), have been noted (Reilly et al., 2019). While there are some studies supporting the view of gender gap in mathematics, some studies are suggesting the opposite, especially in the current years. Fahle

(2016) indicated that gender gap begins in kindergarten and is present until the eighth grade. On the other hand, Penner & Paret (2007) suggested that gender differences appear between the periods of middle school and high school.

According to the researchers, a gender disparity is present in learning styles, concerning mathematics (Geist & King, 2008). Additionally, there are some other factors present that affect math performances between genders such as classroom environment, learning style, and teaching style. For instance, it was revealed that it may not be efficient to expect both genders to benefit equally from the same instruction (Geist & King, 2008). Improving the syllabus and teaching style to make them developmentally applicable, inclusive, and gender sensitive would be beneficial for students to achieve various outcomes (Geist & King, 2008). Regarding classroom environment, single-gender classrooms were thought to be beneficial for students to achieve their potential (Bacchus, 2015). Smith (2011) conducted a study on academic performances including math between single-gender and co-educational classrooms. The results showed that 5^{th} graders who were in the single-gender classroom exhibited higher performances compared to children who were in the co-educational classroom. The study proved that single-gender classes affected children's math performance positively. Sutton (2009) explains this situation as children getting an opportunity to receive a gender specialized education which gives an emphasis on diverse learning styles between genders. Therefore, researchers discussed creating an opportunity for females by creating an environment that promotes concentration on critical thinking and effort skills, and by supporting their self-confidence through the elimination of stereotypes. In addition, throughout the national tests involving mathematics branch the gender gap is present with a male advantage and classroom environment is one of the reasons (Niederle & Vesterlund, 2010). The researchers discussed that the reason is not due to genetic aspects between females and males but rather the impact of competitive test atmosphere on females. In face of competition, females result in lower achievement scores but there is no such effect in the non-competitive atmosphere (Niederle & Vesterlund, 2010).

In comparison, Bacchus (2015) conducted a study including both single-gender and co-educational classrooms to investigate differences in Algebra test performance among 9th graders and found no evidence favouring single-gender classes over co-educational classrooms in terms of math performances. Conversely, children who were in the co-educational classes achieved better results in Algebra compared to children in single-gender classes.

Furthermore, Saleh & Rahman (2016) conducted a study on students' Algebra scores concerning gender and types of schools. The study included three types of schools such as national schools, religious schools, and boarding schools. It was found that children that are in boarding school achieve higher score in Algebra tests compared to children in religious and national schools. Researchers showed that different types of learning environments could affect learning styles. The reason could be explained as boarding schools select students that achieves higher performances. Moreover, this study discussed that there are more learning style differences between females and males in terms of gender. That is, females are more patient during extended hours of classes compared to males due to hormonal reasons. Specifically, males have high levels of testosterone, which makes them less concentrated and more hyperactive. In addition, it was revealed that females displayed higher scores when they practice and study with the exercises that are given by the teacher whereas males were generally better to comprehend Algebra subjects (Saleh & Rahman, 2016). However, the results demonstrated that females displayed higher scores in Algebra tests compared to males Moreover, Musa et al., (2016) said that males were more advantageous when learning math and other lessons through a goal-oriented strategy more than females whereas there was no difference between the scores of math. This small difference may underscore the importance of motivation because when children are motivated, they tend to be more goal-oriented, which helps them achieve academically (Musa et al., 2016). Besides, males and females exhibited different results in terms of mathematical internalization and application into other areas of mathematics, involving mathematics content, in related areas, and in daily tasks (Sari et al., 2020). Females and males differed within mathematical internalization and application of competencies with a male advantage in general. It was explained that females showed higher performances in mathematics however they could not associate the contents to other related things.

According to Penner & Paret (2007), males tend to either achieve the best results or be at the bottom compared to their female classmates in the kindergarten years. Differently, males performed evenly or better compared to their female classmates in the 3rd grade. On the other hand, there are some studies find the opposite result or no result at all. Robinson & Lubienski (2011) revealed a study that is constituted standardized assessment test and teacher ratings in mathematics for kindergarten, middle, and high school. There was no gender gap between females and males in terms of math achievement tests during kindergarten however teacher ratings were higher in favour of females. In the first grade, males achieved higher scores in math achievement tests than females; however, teacher ratings were still in favour of females. Among first and third grades, the gender gap was still present with a male advantage over females. However, the attitudes of teachers remained still by rating females higher (Robinson & Lubienski, 2011). McMillian et al. (2011) revealed results indicating no gender gap in mathematics assessments between males and females aged 8 to 12. Moreover, Reardon et al. (2016) revealed that there was no significant difference between females and males in terms of standardized mathematics assessment. However, another study detected a male advantage in mathematics achievement in high school (Brunner et al., 2008). Researchers considered general cognitive ability and specific mathematical ability of 9th graders of both genders and they found that males outperformed females in mathematical achievement. In accordance with it, Boaler et al. (2011) found that math performances are even for both genders, but males tend to move forward to a high-level in mathematics more than females. Conversely, in a study that is conducted in Punjab for students that are in higher secondary schools or college, females achieved higher scores in mathematics compared to males in mathematics assessment (Malik, 2016). In addition, Musa et al., (2016) revealed different and engaging results on gender difference in math between senior secondary school students in Nigeria. The study showed that males achieved higher scores in English Language and general academic achievement compared to females, but not for mathematics achievement. There are controversial results on mathematics performance regarding gender, some suggesting a male advantage, while others suggesting a female advantage.

CHAPTER 2

Differences in Mathematical Acquisition

2.1 Math content

The content of mathematics assessments could be essential for understanding the possible reasons behind the discrepancies between females and males. As seen in the literature review on gender differences in mathematics, several studies indicate that the results vary depending on the specific type of mathematical content being examined. Gender differences in mathematics assessment may be influenced by the math content (Reardon et al., 2016). Besides, researchers revealed that, there are no distinct differences between females and males in terms of mathematics scores when a standardized mathematics test has been utilized.

To start with calculation skills as a math content, researchers could not find any significant result in relation to gender of students (Tobia et al., 2015). However, they underlined the importance of calculation skills as math content on upcoming mathematical performance. It was found that preliminary calculation skills in preschool period would be beneficial to anticipate later mathematical ability in the upcoming school years (Tobia et al., 2015). To anticipate preliminary number competence, size seriation task yielded important results. The study included two preliminary calculation tests such as addition and subtraction as well. However, as indicated there was not a notable gender gap between students. Accordingly, Plaisted et al. (2011) discussed that males performed higher on Raven's Matrices scores than females. Similarly, Hughes-Isley (2015) found similar scores in pre-college Algebra tests favouring males compared to females. However, there were no such effect in other domains of mathematics. Specifically, the study consisted tests of mathematical ability including analytic 1 and 2. Males achieved higher scores on math with Analytic 2 than math with Analytic 1 compared to females (Hughes-Isley, 2015).

In addition, Shen et al. (2016) published a study on gender differences in mathematics assessment including arithmetic. The study conducted on students who were in the first year of school from the United States, Russia, and Taiwan. Students were asked to complete simple and complex addition problems and to indicate which strategy they implemented. There occurred no significant difference between genders in terms of simple problems, however; there was a significant difference between genders in favour of males in terms of complex questions for students from the United States and Russia, based on the strategy they use. This study helped to understand how math content could influence the results in relation to gender. Conversely, Wei et al. (2012) said that females are generally getting higher scores on arithmetic in a word-rhyming task. Students that are aged between 8-11 from Beijing assessed with arithmetic tasks such as simple subtraction and complex multiplication while adjusting for word-rhyming task. However, the results indicated that after adjusting for the word-rhyming task, there was no significant gender difference for arithmetic. Nevertheless, partly because females demonstrated higher performance in language processing, they found it easier to achieve higher scores in arithmetic when the word-rhyming was not controlled for. Similarly, Carr et al. (2001) conducted another study on the usage of strategy in arithmetic to reveal gender differences. The study included two conditions: a free-choice condition, in which participants could use whatever strategy they wanted to solve the problems, or a game-condition, in they were instructed to use one strategy for each problem. The results indicated that in the free-choice condition, females used manipulatives as a strategy whereas males used retrieval. In the game condition, both females and males utilized manipulatives to solve the problems with males being more variable compared to females. Even though, diverse results are present for calculation and arithmetic tasks suggesting an effect or no effect at all for gender, the result may seem to favour more males than females.

Furthermore, Erdogan et al. (2011) conducted a study that focused on another particular math content which is geometry. Researchers highlighted the significance of personal differences in academical performance, with gender being one of the most important components associated with learning style. Personal variances can impact the way students internalize knowledge across different domains, such as mathematics and geometry (Erdogan et al., 2011). The results showed that females achieved higher self-efficacy beliefs and scores in Geometry compared to male students. However, in another study, it was found that males outperform females on spatial and abstraction abilities in geometry tests, attributed to their more complex spatial skills (Fitriani & Nurfauziah, 2019). Researchers indicated that males are better in the tasks involving critical and complex thinking. On the other hand, females are better to associate the interpretation of the problem. However, spatial abilities need to be elaborated separately than problem-solving geometry tasks due to its complex and diverse nature than mathematics. There are several studies suggests that visuo-spatial skills have a different viewpoint than mathematics (Fitriani & Nurfauziah, 2019). Thus, the results may have been related with visuo-spatial skills which should be elaborated separately than mathematics content. On the other hand, its high complexity may support the idea that males achieve higher scores in tests that acquire higher complexity. Additionally, Erdogan (2008) revealed a study on the elementary school teachers' geometry abilities with van Hiele theory. This theory consists of five different levels when performing on a geometry question and they are visualization, analysis, ordering, deduction, rigor. They found that there were no gender differences in thinking styles, whereas there appeared gender differences in reasoning levels in favour of males. Thus, there might be some gender differences in geometry in terms of reasoning with a male advantage.

In addition, Thomas et al. (2022) published a study that examines the gender gap in numeracy tests within the National Test from 2008 to 2016 in Australia. The numeracy test included comprehension, fluency, problem solving, and reasoning in terms of number and algebra, measurement and geometry, and statistics and probability for the Years 3, 5, 7, and 9. Males displayed higher results in terms of numeracy test performance compared to males in all years. On the other hand, Begum et al. (2021) showed different results regarding numeracy between genders. There were no significant discrepancies between females and males based on the numeracy test that they undertook.

Finally, TIMSS study that has been conducted in 2003 demonstrated different outcomes regarding math achievement. Researchers discussed that the results of mathematics assessments can vary based on the education level of every country (Ayalon et al., 2013). Indeed, they found that the gender difference disappears in the presence of a national tests, meaning that males do not outperform females. Another study, Mann et al. (2016) showed that national tests environments are helpful to close the gender gap by promoting the achievement of females.

2. 2 Previous studies

Previous studies were extracted to obtain knowledge on the topic of interest and gain insight from other meta-analysis studies. These meta-analyses suggested various and notable results that worth discussing in terms of new trends in mathematics performances and cross-cultural differences when considering gender differences in mathematics. Lindberg and colleagues (2010) found that the gender gap was not present in mathematics performance currently whereas Else-Quest (2010) indicated that there were some differences present regarding gender-equality according to nations. To start with, Lindberg et al. (2010) has revealed a meta-analysis study on gender differences in mathematics performance within emphasis on new trends. They carried out a meta-analysis on 242 articles released between 1990 and 2007, involving a sample size of 1,286,350 individuals. The study found no significant gender difference in mathematics, with a Cohen's *d* score of .05 suggesting a very small gender difference in favour of males. Later, they conducted another examination on students from United States for the last 20 years from the large databases such as The National Longitudinal Survey of Youth – 1997 (NLSY), The National Educational Longitudinal Study (NELS88), The Longitudinal Study of American Youth (LSAY), and the National Assessment of Educational Progress (NAEP). Similarly, there did not occurred a certain significant difference with an effect size that is ranging between -0.15 to +0.22.

Firstly, they highlighted the significance of gender stereotypes within mathematical performances, suggesting that these stereotypes create obstacle for women in fields such as science, technology, engineering, and mathematics (STEM). According to the researchers, there are some notable stereotypes regarding male superiority in mathematics that encompasses children, adolescents, families and even teachers. These stereotypes can impact female students' self-competency and self-efficacy beliefs and affect their achievement scores.

Moreover, it was also indicated that previous literature has shown diverse results, some supporting male superiority or other supporting female superiority, based on some other facilitator variables. One of these variables were age and difficulty level of the test separated as computation (the minimum level of difficulty), understanding the mathematical contents (medium level), complex problem-solving (maximum level of difficulty) (Hyde et al., 1990a cited by Lindberg et al., 2010). In accordance, the results showed that females achieved higher scores in computation (minimum level) compared to males in elementary and middle school whereas there occurred no discrepancies in high school. Similarly, there was no significant gender difference in understanding the mathematical contents in any school period. Lastly, males achieved higher scores in complex problem-solving than females in high school, but no such effect emerged in elementary and middle school. These results should be taken into consider because complex problem-solving is essential for future career in mathematics and STEM. However, they also highlighted that these gender differences start to decrease over the years as female students become more involved in areas such as mathematics and STEM. Furthermore, the gender gap in terms of complex problem solving has begun to disappear. Another point was that males tend to be the first or the last in mathematics performance and this suggests a higher male variability. That is why researchers aimed to conduct a new meta-analysis to further search the gender differences in mathematics by taking into consider variability.

The first study examined whether depth of knowledge, age, cultural differences, and the year of publication moderated the gender difference in mathematics performance. The results showed that there was no significant gender difference; however, some moderator variables were significant. Problem type (multiple choice, short answer, open ended questions) demonstrated significant results indicating that males outperformed females in the tests with multiple choice whereas females outperform males in the presence of an open-ended and short answer test. Another significant moderator was age suggesting a gender difference in high school, but not in elementary and middle school. There was a decline in the gender gap also during college period and adulthood.

Researchers conducted another study which focuses on large datasets from United States for the last 20 years. These national large datasets are the NLSY-1997, the NELS-88, the LSAY, and the NAEP. The NLYS – 1997 extracted data between 1997 to 2002 and included a mathematics performance test called PIAT-R involving multiple choice items on three different topics: foundations (number size, and shape discretion), basic facts (addition, subtraction, multiplication, division), applications (algebra, geometry, fractions, word problems, and numerical relationships). This test included some specific measurements, such as whether students achieved a successful score on questions that are suitable for their ages. The next step was to increase the difficulty level if they were successful, or not to increase and even decrease it if they were not successful. The NELS88 extracted data from the years 1988, 1990, and 1992 for middle and high school, as it was a longitudinal study. The mathematics achievement test included multiple choice questions involving arithmetic, algebra, geometry, data and probability, and advanced topics. The students were challenged according to their mathematics skills in the upcoming years with an emphasis on skill/knowledge, comprehension, and problem solving. The LSAY retrieved data from 1987 to 1992 for middle and high school students, which included a mathematics test covering geometry, measurement, data analysis, algebra and simple operations in a multiple-choice format. Across all, NAEP was a large database, but it was not longitudinal like the others. The NAEP database encompassed two measurements: the long-term trend assessment and the main assessment. The long-term trend assessment examined scores from 1992 to 2004 in every four years for students in elementary, middle and high school. The math assessment included questions such as multiple-choice and short answers on number operations, measurement, algebra, and geometry. Conversely, the main assessment retrieved data from 4th and the 8th graders every two years from 1990 to 2007, while data from 12th graders was retrieved only in 1990 to 2007. The mathematics measurement included multiple choice, short-answered, and long constructed response items on mathematics abilities and data and probability. Interestingly, the NAEP assessment encompassed item-response theory, which measured students based on their comprehension of mathematics instead of simply whether their answers were correct or incorrect. The study examined whether age, publication year, percentage of each type of problem (number sense, algebra, geometry, measurement), percentage of problems in each type of format (multiple choice, short answer, open ended), ethnic diversity, and variability moderated the gender difference in general.

The overall results indicated not significant differences, with a Cohen's *d* of 0.07. However, since the results demonstrated higher levels of heterogeneity, researchers decided to introduce moderators to better understand the reasons for this variability. One of the moderators, the problem type, demonstrated a small significant outcome for multiple choice and short answer items, while showing large significant outcome for short answer items. However, regarding NAEP, the results showed that males outperform females in multiple-choice items, whereas females outperform males in short-answer and open-ended questions. In summary, there was a significant effect of gender on the problem type in the NAEP assessments. The other moderator, mathematical content, which females achieved higher scores among algebra questions more than males, whereas males achieved better scores among measurement questions. In terms of depth of knowledge, questions that are in the Levels 3 and 4 which includes complex problem-solving questions demonstrated significant differences in terms of gender. That is, males achieved higher scores in the tests of Level 3 and 4 more than females whereas the same effect was not observed in Study 1.

Lindberg et al. (2010) contributed to the literature by stating that there is currently no considerable gender difference. Specifically, the gender gap started to decrease, and the gender discrepancies were no longer clearly visible.

Furthermore, higher attention was given to another meta-analysis by Else-Quest et al. (2010), which published an article on cross-national patterns of gender differences. According to their literature review, it was possible to find gender differences in some countries, but not in others. Additionally, according to the gender stratification hypothesis, gender differences may derive from the cultural environment that females are surrounded by. The theoretical structure underlines this hypothesis, which suggests that in some countries, social hierarchy is present in terms of gender. Thus, this affects females' attitudes towards math, causing them to display more negative affect, which in turn impacts their math performances. As Lindberg et al. (2010) stated, in this study researchers also underlined the significance of small number of women in STEM in most of the countries as well. This meta-analysis study involved two large international datasets, the 2003 Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) encompassing a total of 493,495 students aged between 14-16 from 69 countries. The scope of this search was to investigate gender differences in mathematics achievement, attitudes, and affect and their associations with gender equality internationally. Researchers revealed that recent literature is mostly based on the samples from the United States, so it might be useful to investigate these differences internationally with another study, as there may be different results for every nation. Throughout their literature review, there were several studies that indicate no significant gender differences currently. This is in line with the gender similarities hypothesis, which suggests that despite some several psychological differences, males and females are alike. However, despite gender similarities in mathematics achievement there are still stereotypes about women underachieve in mathematics. Thus, this meta-analysis study aimed to further investigate these differences in detail.

To get into detail into the gender differences, attitudes, and affect in mathematics it is essential to learn about TIMSS and PISA. TIMSS is an international database that investigates mathematics and science learning of children in 8th grade on every four years. The execution held by the International Association for the Evaluation of International Achievement (IEA) together with Statistics Canada and Educational Testing Service. Similarly, PISA is also an international study implemented by the Organisation for Economic Co-operation and Development (OECD), examining mathematics, reading, science, and problem-solving literacy of 15-year-old children every three years. These two large international datasets serve different purposes. TIMMS focuses more on the syllabus and assesses what has been taught in schools, along with teacher and school related aspects. On the other hand, PISA pays attention to mathematical literacy, helping in understanding the practical application of mathematics in real life. It helps individuals to interpret, employ, and interact with mathematics in their daily lives. Since PISA is an assessment that is more applied, it may be more difficult to internalize mathematics and apply it when considering PISA. PISA seems more complex since it requires more cognitive and practical processes than TIMSS. According to what Lindberg et al. (2010) had discussed that in one study males achieved higher scores in the presence of a complex problem-solving test compared to females. Thus, PISA might be more useful to interpret gender differences due to its challenges and can illustrate the male variability more clearly. On the other hand, TIMSS serves more comprehensive and varied sample by involving under-developed nations in contrast to PISA. Nevertheless, both TIMSS and PISA involve more developed countries, and this creates a challenge by generating a non-representative sample. In general, both TIMSS and PISA might be useful to understand the variability due to their differences.

In this study, the TIMMS 2003 carried out in 46 nations encompassing 219,612 students and involved five subjects to assess mathematics achievement: number, algebra, measurement, geometry, and data. The number subject involves whole numbers, fractions, decimals, integers, ratios, proportion and percentages. The algebra domain includes understanding of patterns, algebraic expressions, equations, formulas, and relationships. Measurement subject encompasses the content of attributes and units and tools, techniques, and formulas. The geometry content examines knowledge of lines and angles, two and three-dimensional shapes, congruence, similarity, locations, spatial relationships, symmetry, and formulas. Lastly, data subject involves, topics of data collection and organization, data representation, data interpretation, uncertainty and probability. The next step was to examine attitudes and affect. To understand these, students were instructed to use two scales: one for self-confidence and the other for their evaluation of mathematics.

In contrast, PISA 2003 extracted data from 41 nations involving 273, 883 students by encompassing four subjects to examine mathematics performances: quantity, space/shape, change/relationships, and uncertainty. The quantity domain examines the understanding of numeric phenomena, quantitative relationships, and patterns, and it is like the number domain of TIMSS. The space/shape content includes understanding of spatial and geometric phenomena, and relationships, and it corresponds to the geometry content in TIMSS. The change/relationships domain assesses understanding of mathematical manifestations of change, functional relationships, and dependency across variables, it is comparable with Algebra content in TIMSS. Lastly, the domain of Uncertainty examines understanding of probabilities and statistics, and it resembles the data domain of TIMSS. To assess attitudes and affect, the PISA 2003 carried out five different scales such as extrinsic motivation, intrinsic motivation, anxiety, self-concept, self-efficacy.

To measure gender equality and understand the gender stratification theory, researchers utilized recent indexes from the literature reviews. These clarified gender differences in areas such as economic, educational, and political occasions. Composite benchmarks consisted of Gender Development Index (GDI), Gender Empowerment Measure (GEM), Gender Equality Index (GEI), and Standardized Index of Gender Equality Index (SIGE). Domain-specific benchmarks were also utilized due to the constraints of composite benchmarks. These domain specific indexes involved education, political, and economic occasions in gender equality. In summary, the composite and domain-specific indexes were moderators for the gender differences in mathematics performances, attitudes and affect across nations.

The results for TIMSS math data showed that, there was no significant difference among genders in mathematics performances with an effect size of $d = -0.01$. An interesting result was that females achieved higher scores in Algebra compared to males. In general, there was no male variability depending on the complexity level of the TIMSS math assessment. For attitudes and affect, self-confidence test demonstrated an effect size of $d = 0.15$, meaning that males slightly displayed higher self-confidence than females. For valuing math, the effect size recorded as $d = 0.10$, meaning a slight significant result in favour of males. In general, self-confidence and valuing math were slightly significant.

The results for PISA math assessment, a slight but significant difference was observed in mathematics performances between genders, with an effect size of $d = 0.11$ in favour of males. For the attitudes and affect results; extrinsic motivation, intrinsic motivation, anxiety, self-efficacy, and self-concept showed small effect sizes. However, there were some several differences such as males displayed more positive attitude towards math than females ($d = 0.21$ to $d = 0.33$). Females had higher anxiety scores compared to males with a $d = -0.27$. In summary, males displayed higher extrinsic and intrinsic motivation, self-report, and self-efficacy but lower anxiety levels.

Regarding moderators, composite indexes was examined for gender equality to understand the gender differences more clearly in mathematics, attitudes, and affect. The results found no significant evidence of gender equality as a moderator while interpreting gender differences for TIMSS assessment. For PISA assessment, results revealed different results indicating that four composite indexes demonstrated significant results as a moderator in mathematics achievement in favour of males. For attitudes and affect, the GEM showed significant results in all but one case of attitudes and affect scales. However, results surprisingly showed that males scored higher scores in extrinsic motivation, intrinsic motivation, self-concept, self-efficacy in the countries that have high gender equality whereas females scored higher in anxiety. In terms of domain-specific indexes for gender equality, similarly, aim was to examine the gender differences in mathematics, attitudes, and affect. In summary, gender equality is helpful to interpret the gender differences in mathematics, attitudes, and affect across nations. For both PISA and TIMSS, proportion of women in research roles, explained the gender differences, meaning that in nations where women participated in research roles there was nearly no gender difference in self-confidence, valuing math, intrinsic and extrinsic motivation, math anxiety, self-efficacy, and self-concept.

As a summary, the study proved that males and females do not show large differences in their mathematics performances, but males show higher positive affect and attitude with a variability among nations. Additionally, women's condition and well-being affected the results in mathematics between genders internationally, depending on the nation.

CHAPTER 3

METHODOLOGY

The statistical technique utilized in this research is meta-analysis, which comprehensively synthesizes and aggregates outcomes of more than one scientific study. The purpose is to conduct a broad investigation of the field of interest for researchers, examining and gathering all necessary data. It is identified as a synthesis of all the pertinent information and data, which may enhance the reliability and accuracy of inferences compared to individual studies. As it allows researchers to understand what causes the difference between the outcomes of the studies, also it might be beneficial to comprehend how substantial the observed impact is and how consistent the outcomes are across studies.

This study aims to perform a comprehensive review of gender differences in mathematics to compare and discuss the results. In the selection and coding phase, the PRISMA guideline (Page et al., 2021) has been chosen to consider various studies that are of interest. Furthermore, it has been helpful to follow the checklist of the PRISMA 2020 statement for the inclusion of studies that are accurate, reliable, and consistent.

3.1 Aims and Eligibility Criteria

The meta-analysis study that we had conducted contributes to broader collaborative research coordinated by the University of Genova and University of Padova. This research aims to explore potential gender differences in mathematics. To understand these differences between genders, extensive research needs to be done throughout the relevant existing literature, and studies that meet the specified criteria should be selected as stated in the PRISMA checklist.

The present work aimed to consider the data from complete texts that had published in the years 2011 and 2016 with the objective to address these research questions:

- Do boys and girls have a difference on their mathematics score?
- Does the maths content moderate this difference?

By the PRISMA guideline, the implementation of the meta-analysis constituted different steps: conducting a comprehensive literature search among studies that are of interest, evaluation of the abstracts, and evaluation and final coding of the relevant complete texts.

3.1.1. Literature Search

In this part, it was considered to select as many distinct databases as possible to access to all the studies that were of interest. To be able to perform a comprehensive literature search that also evades publication bias, apart from published studies, theses and book chapters (that is called grey literature) have been retrieved from PsycINFO (EBSCO), Scopus, Medline (Ovid), Web of Knowledge, and PubMed since it is substantial to observe all the existing literature to have thorough, exhaustive, and objective results. The literature search was implemented using these keywords:

1) (gender OR sex OR male* OR female* OR girl* OR boy*) AND (math* OR geometry OR arith*) AND (performance OR achievement OR abilit*) AND (child* OR adolesc* OR student*) NOT (“mathematical model”);

2) (gender OR sex OR male* OR female* OR girl* OR boy*) AND (math* OR calculus OR algebra OR geometry OR arith*) AND (performance OR achievement OR abilit*) AND (child* OR adolesc* OR student*) NOT (“mathematical model”).

The composition of each word was adjusted to the specific characteristics of the website that enables an in-depth search of all the related studies. To consider all the forms of the keywords as listed above, which might be written with a different form or vocabulary, the wildcard "*" (asterisk) has been used. The terms "gender", "sex", "female", "girl", "boy", "child", "adolescent", and "student" were used to pick out studies that involves both genders and all the age groups. Moreover, the terms "math", "geometry", "arith", "performance", "achievement", and "ability" denote the mathematical tasks utilized in the studies. However, the phrase "mathematical model" has been used to reject publications related to math. This process facilitated as many beneficial studies as possible to include in the literature search. The existing search comprised of 18,938 records from five different databases between January 2010 and December 2022. Specifically, 1,481 studies in the existing literature for the years 2011 and 2016 are considered in the present study (see Figure XX).

3.1.2 Title and Abstract Screening

At the beginning of the process, the duplicate extraction was implemented using the EndNote application and Rayyan software; once detected, the duplicates were removed, leaving a total of 970 records. Since, it was not sufficient for an article to merely contain all the specified phrases, it was important to read the abstracts to determine if the article corresponds to our research interest. Herein, 369 abstracts were rejected based on the selection criteria. Double-screening process was carried out on the 20% of the abstracts to prevent any overlook and omission. Discussions had been held between Lorenzo Esposito and Sena Nur Kaya to resolve any discrepancies or inconsistencies and reach an agreement.

As agreed, upon before the literature search, the inclusion criteria for the initial screening phase are as follows:

1. The existence of an assessment in mathematics or a clear statement that mathematics is being assessed;
2. The presence of both genders (exclusion of single-gender studies);

3. A sample consisting of typically developing individuals older than 4 years old.

Consequently, 601 theses and articles met the inclusion criteria after the abstract screening process.

3.1.3. Full-text Analysis

Henceforth, the further step was to conduct a full-text analysis for the articles and theses that met the inclusion criteria in the previous selection phase. The objective was to screen the complete text, not just abstracts, in order to determine whether they met the criteria. Moreover, a special attention was paid to mathematical assessment and sample population apart from the inclusion criteria mentioned before.

To begin, it was crucial to understand how a mathematical assessment was conducted. In this regard, scores taken from teacher-administered tests or self-reported grades were not taken into consider. In addition, mathematical assessments which included geometry were another aspect to consider, as they involve spatial tasks that should be excluded due to their challenging evaluation, which differs from mathematics assessment. However, emphasis was placed on articles that involve problem-solving in geometry and geometrical concepts. In relation to the sample population, concern was the presence of typically developing individuals. In studies involving atypical populations, only the typical group was considered; in intervention studies all waves were considered; while for studies including experimental manipulations, only the control group was considered.

The information of the articles was coded into Excel as follows:

1. The article ID (a progressive number to identify the entry in the dataset);
2. The publication year;
3. The author;
4. The title;

5. The full-text file (1: retrieved, 0: not retrieved);
6. Inclusion (1: included, 0: excluded);
7. Data (1: Data available in the article, 0: data not available);
8. Details on exclusion reason;
9. Email to contact the author;
10. The information requested for the author;

3.1.4. Coding

Following the examination of abstract and the full-text screening, the last stage of the study has started which is coding phase. During this phase, 5.13% of the studies remained (as indicated in the initial records) resulting in the validation of 76 articles and theses. This coding phase was essential for constructing a platform that encompasses all the necessary information from the studies within the area of interest. The information that is needed for conducting analyses was determined by the researchers based on the requirements. Despite some articles lacking all the required information, they were still coded due to their inclusion of outcomes of the mathematical assessment and presence of both genders typically developed. The remained information, coded into the Excel, as follows:

1. The article ID;
2. The ID sample (a nested ID to identify different samples inside the same study);
3. The authors;
4. Notes;
5. Publication type (1: journal article, 0: thesis, 2: other);
6. Data (1: yes, 2: data but a sample and/or measure is missing);
7. Survey (1: yes, 0: no);
8. Survey type, the name of the survey;

9. Survey year, the administration year of the survey;

10. Male sample size;

11. Female sample size;

12. Task name;

13. Standardized test (1: standardized, 0: not standardized);

14. Task description (content type);

15. Computerized (1: computerized task, 0: not computerized;)

16. Male mean;

17. Male standard deviation;

18. Female mean;

19. Female standard deviation;

20. Index name (accuracy, reaction time, error);

21. School level (kindergarten, primary, middle, high, university or adulthood);

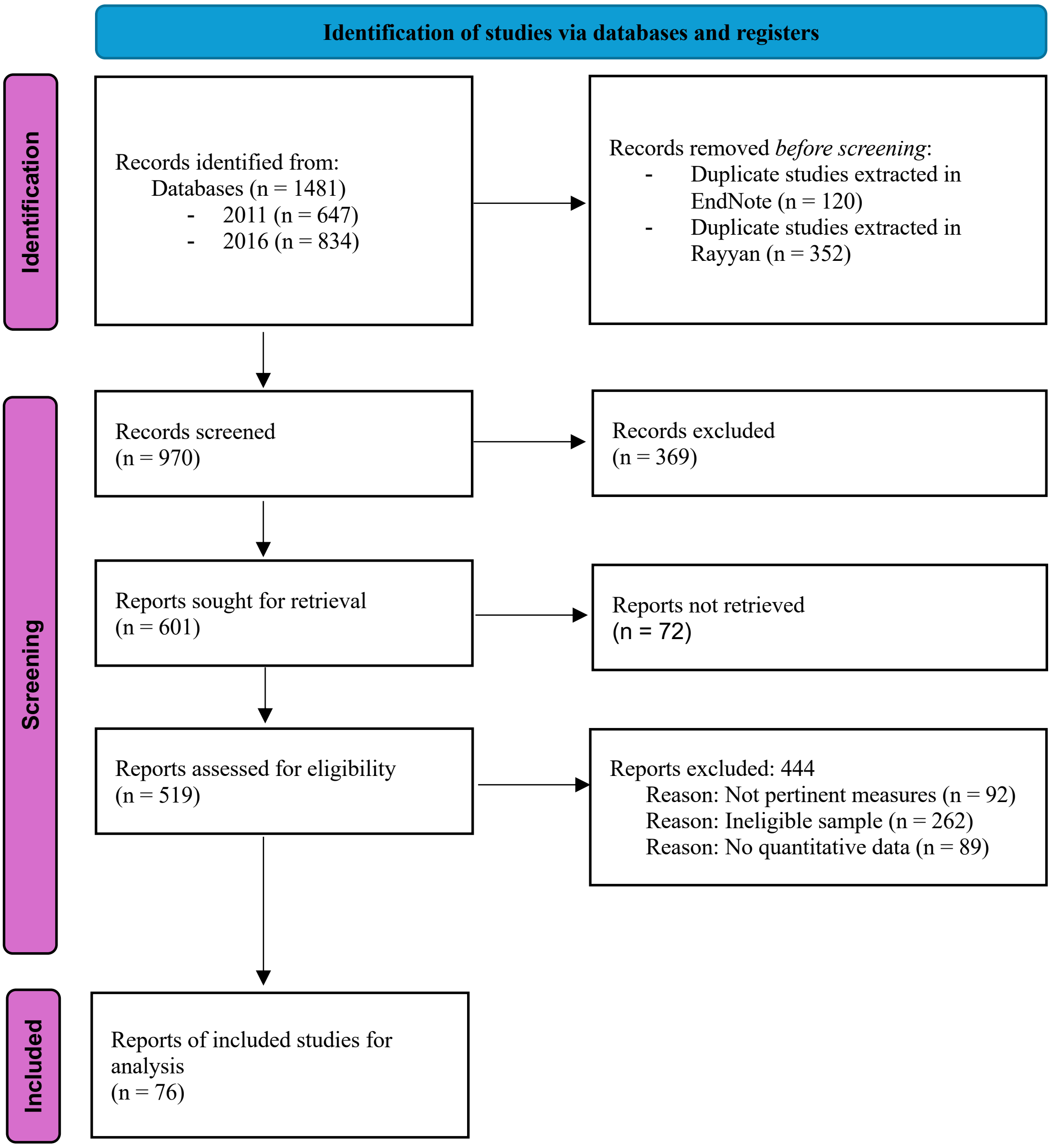
22. Country;

23. Age range sample;

3.2 Analytical Strategy

As a measure of effect size Cohen's *d* was utilized. It quantifies the standardized difference between two groups. Groups were assigned codes of 0 and 1, with 0 representing male samples (M = 0) and 1 representing female samples (F = 1). In instances where a positive Cohen's d is observed, there is a disparity favouring males over females; conversely, when a negative Cohen's d is observed, there is a disparity favouring females over males. For each study, it was opted to utilize only one Cohen's d even with in cases there might be more than one Cohen's d values available. In such cases, the average score was calculated. Some of the studies that were coded were of a single sample, comprising 55%, whereas 44.59% of the papers contained multiple samples and outcomes. The papers that contained multiple samples and outcomes were coded independently from one another. This approach was also implemented for the studies that contains diverse assessments. After computing Cohen's *d*, its variance was calculated, and multivariate meta-analysis models were fitted by using R-Studio package "metafor".

Figure 1. PRISMA 2020 flow diagram for systematic reviews that containing databases and registers



CHAPTER 4

RESULTS

Analyses have been conducted to determine whether there is a difference between the mathematics assessments of females and males, with emphasis placed on whether math content plays a role in the results or not. This meta-analysis contained 76 articles that are published in the years 2011 and 2016, incorporating 76 samples and encompassing a total of 508,429 females and 527,857 males. Descriptive statistics are reported in Table 1.

Table 1

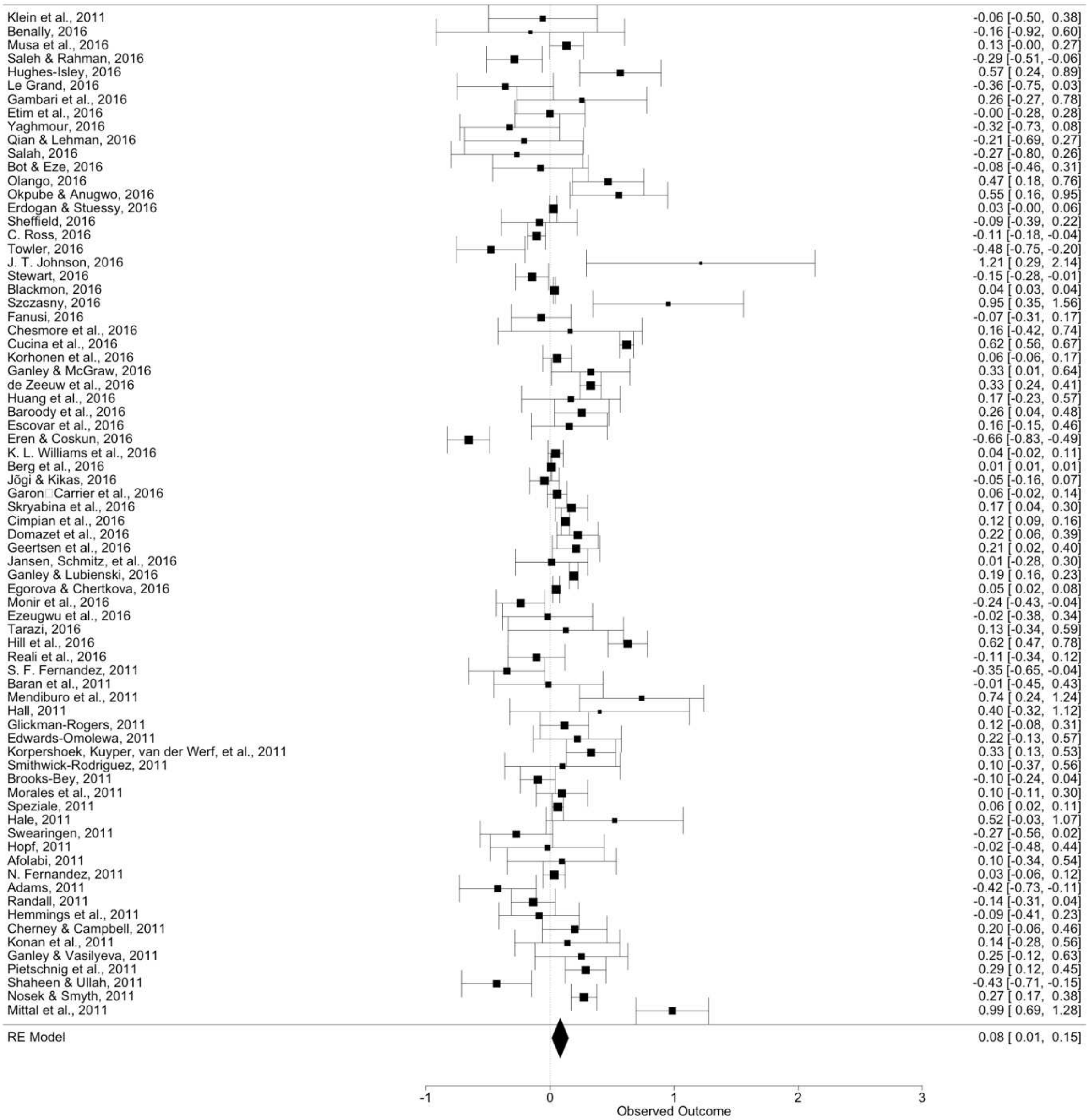
Descriptive statistics by math content

Math Content	*n* studies	*n* effect sizes	*n* males	*n* females	*d*	CI%95	*SE*
Arithmetic	7	12	3360	3386	0.372	[0.085, 0.658]	0.146
Broad Mathematics	26	54	394477	374315	0.054	[-0.047, 0.155]	0.052
National Test	31	53	128742	128984	0.020	[-0.061, 0.100]	0.041
Advanced Maths	7	14	365	443	0.086	[-0.203, 0.375]	0.147
Basic Numeracy	3	4	860	1253	0.144	[-0.029, 0.316]	0.088
Geometry	2	6	53	48	0.041	[-0.272, 0.353]	0.159

4.1 Forest Plot

Firstly, the forest plot was employed as tool for discerning heterogeneity and inspecting very briefly the results. The forest plot can be interpreted as follows: the studies included in the meta-analysis process are sorted sequentially in the vertical section of the plot. Subsequently, the plot includes squares and lines for every study, where the size of square symbolizes the sample size, the position of square indicating effect size, and the lines symbolizes confidence intervals. To begin with, as the sample size increases for each study, the square size also increases. Conversely, as the sample size decreases, the size of the square decreases accordingly. For instance, the study by Benally (2016) could be illustrated with a smaller square size, representing a smaller sample size compared to other studies, such as Nosek & Smyth (2011). Ensuring a large sample size is substantial for including diverse samples, thereby facilitating a more comprehensive and realistic outcome. Additionally, when examining sample sizes, it is important to consider confidence intervals, which are represented as lines (also referred to as twigs) in the plot. Confidence intervals display the estimation of each study. Extended horizontal lines indicate a less accurate estimate, whereas shorter lines signify a more accurate estimate. The study by Eren & Coskun (2016) demonstrates an example, with shorter lines representing shorter confidence intervals, leading to a more precise estimate compared to studies such as J. T. Johnson (2016). There is a relationship between sample sizes and confidence intervals: as the sample size decreases, confidence interval gets longer and conversely as the sample size increases, confidence interval gets shorter. The scope is to obtain a large sample size with a short confidence interval which will generate more precise and comprehensive results. The position of the square relative to the effect sizes depends on its position relative to the vertical line, which is at the center of the plot. The vertical line itself stands for a null value, suggesting no difference in mathematical performance between females and males. When a study is positioned to the left of the line, it implies a negative effect size, whereas a study positioned to the right implies a positive effect size. As an example, when examining the left side, Towler (2016) exhibits a negative effect, whereas Mittal (2011) implies a positive effect. Besides, this vertical line serves as a reference to determine whether a study is significant or not based on the confidence intervals lines. Specifically, if the confidence interval lines intersect with the vertical line, it means the study lacks statistical significance. On the other hand, if the confidence interval lines do not intersect with the vertical line, the study indicates statistical significance. As an illustration, Fanusi (2016) is not statistically significant because it intersects with the vertical line, whereas Hill et al. (2016) is statistically significant as it does not overlap with the vertical line. Furthermore, the forest plot provides numerical presentation of Cohen's *d*, indicating the effect size. The decision was that a positive effect size suggests males performed better than females, whereas a negative effect size suggests females performed better than males. As seen in the Figure 1, Monir et al. (2016) demonstrated a negative effect size (*d* = -0.24) which suggesting a superiority among females, while Hill et al. (2016) demonstrated a positive effect size (*d* = 0.13) suggesting a superiority among males. The overall estimated effect size could be observed from the diamond shape displayed at the bottom of the plot. It is still important to consider on its position relative to the vertical line, and whether it intersects with it. Diamond shape is positioned at the right side, suggesting male superiority with a positive effect size (*d* = 0.08) and a confidence interval between 95% CI [0.01, 0.15]. It does not intersect the vertical line, suggesting an overall significant effect. The forest plot helps in identifying heterogeneity and outliers. Evidence of heterogeneity is suggested when the general distribution is dispersed across the entire plot. In general, studies are positioned close to the vertical line, which basically lies near to the midpoint of the plot, suggesting a discrete level of heterogeneity.

Figure 2. Illustration of 76 studies in the meta-analysis as a Forest Plot

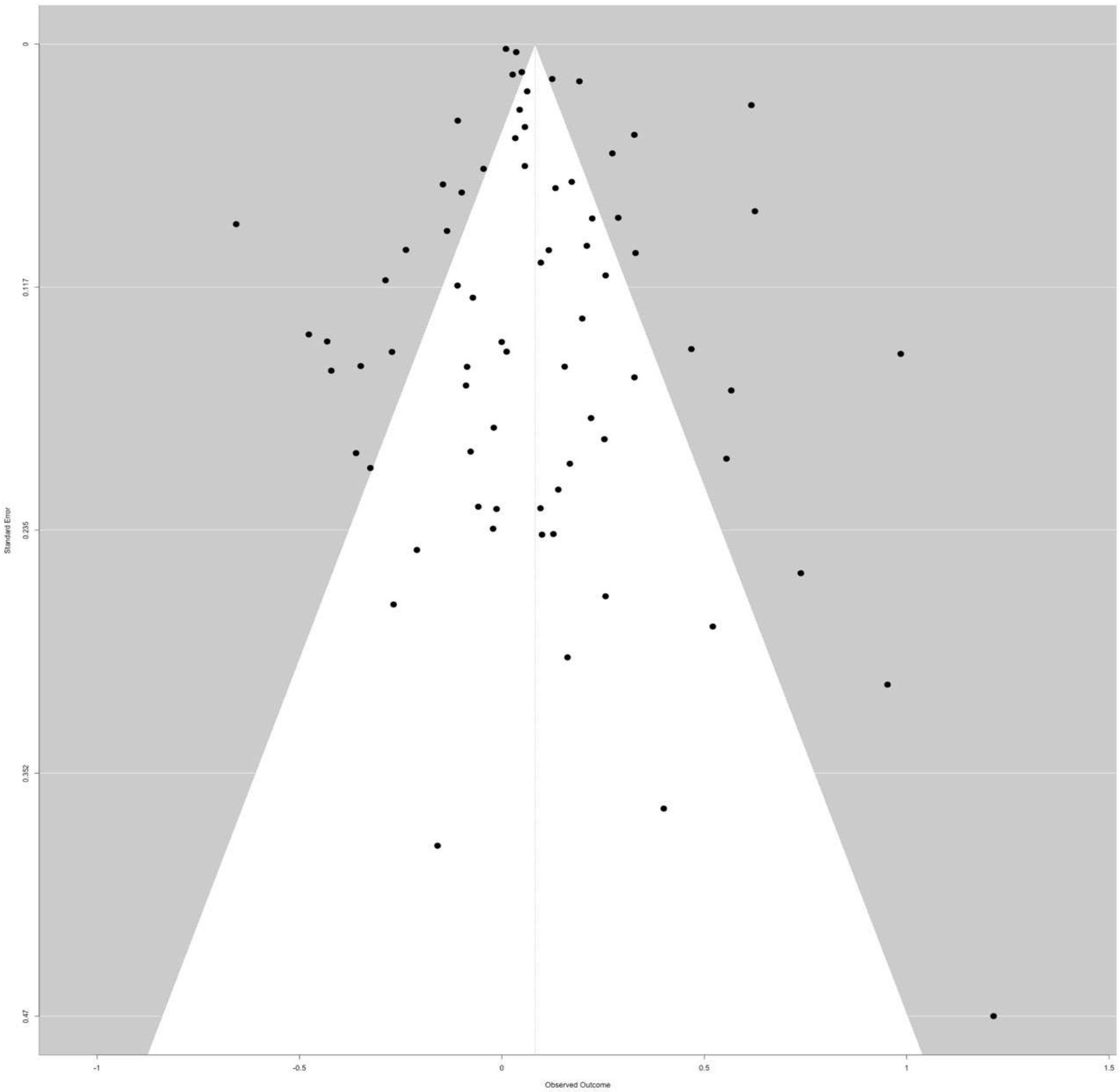


4.2 Funnel Plot

Funnel plot is helpful for examining publication bias by illustrating the general distribution studies across the entire plot. The general distribution is expected to be symmetrical and in the presence of an asymmetrical distribution, publication bias is present. This means, the meta-analysis is measuring only a specific population which does not reflect the whole population, indicating that the meta-analysis is not representative. This meta-analysis study does not suggest publication bias, as a symmetrical distribution of the funnel plot can be observed in Figure 3.

Every point on the funnel plot denotes an individual study within the meta-analysis, positioned based on the outcomes they have reached. In particular, studies positioned at the lower end of the plot are those with outcomes associated with high standard error and a small sample size. Conversely, studies positioned at the top of the plot are those with outcomes associated with low standard error and a large sample size. The funnel plot reveals that the findings of this meta-analysis study follow the expected pattern. To be more precise, the points distributed at the top of the plot indicate studies with large sample sizes and low standard errors.

Figure 3. Illustrates studies published in both 2011 and 2016 using a Funnel Plot



4.3 Analytical results

Subsequently, further analyses have been conducted to elucidate differences in mathematical assessment performances between females and males. As shown in the Figure 1, the forest plot indicates a Cohen's d = 0.08, with a confidence interval 95% CI [0.01, 0.15] suggesting a significant effect size. The SE = 0.03 while the test of heterogeneity yielded significant results, $Q(142)$ = 1896.09, p < .001.

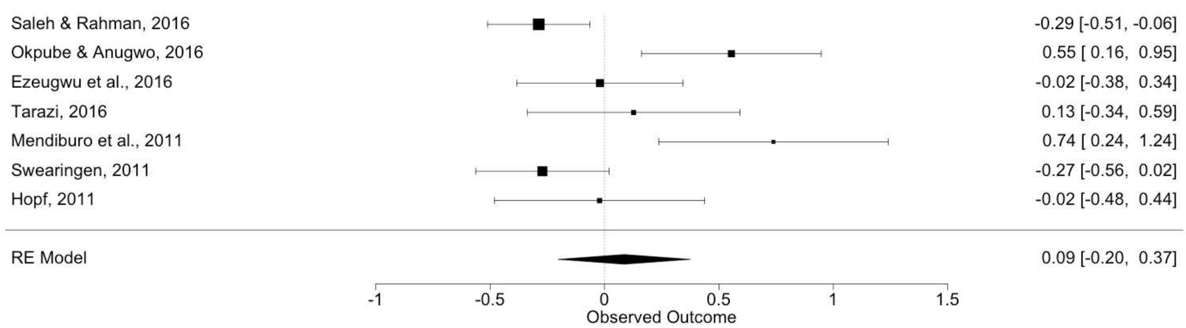
4.4 Math content as a moderator

Therefore, it has been decided to include math content as a moderator in the model, as the model without moderator showed heterogenous results. The reason of adding moderators is that, due to the high heterogeneity, the study might be better explained with an additional variable, as the model without moderators is insufficient to explain all aspects of the outcome.

4.4.1 Advanced math

This mathematical content was selected as the intercept among all other math contents. The meta-analysis did not reveal a significant effect. Estimate is $\beta = 0.03$ (*SE* = 0.11) with a confidence interval ranging from -0.18 to 0.24.

Figure 4. Forest plot for advanced mathematics

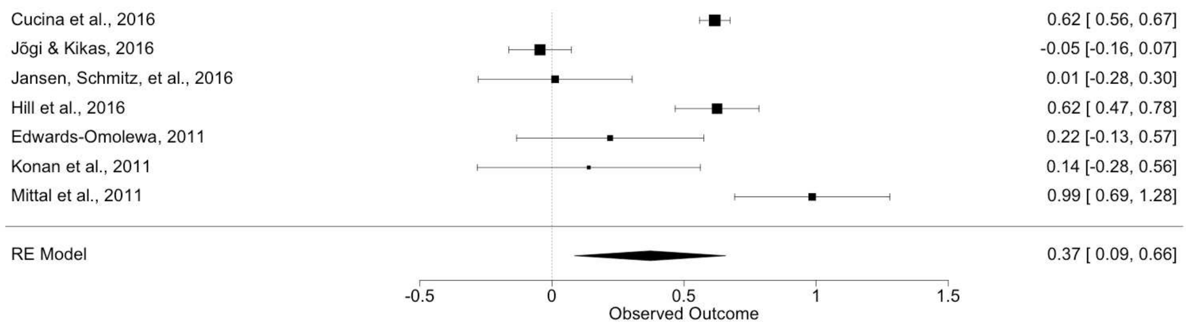


Note. Figure 4 represents the descriptive statistics for this math content.

4.4.2 Arithmetic

The meta-analysis revealed that the arithmetic effect was significant $\beta = 0.33$ (*SE* = 0.14) with a confidence interval ranging from 0.04 to 0.63. The forest plot shows that the diamond shape does not intersect with zero so that means the studies are significant. Additionally, the studies are distributed at the right side of the vertical line meaning, they are positive. Positive studies reveal that male perform better in the mathematical assessments than females, as shown by *d* = 0.37 with a confidence interval 95% CI [0.09, 0.66].

Figure 5. Forest plot for arithmetic

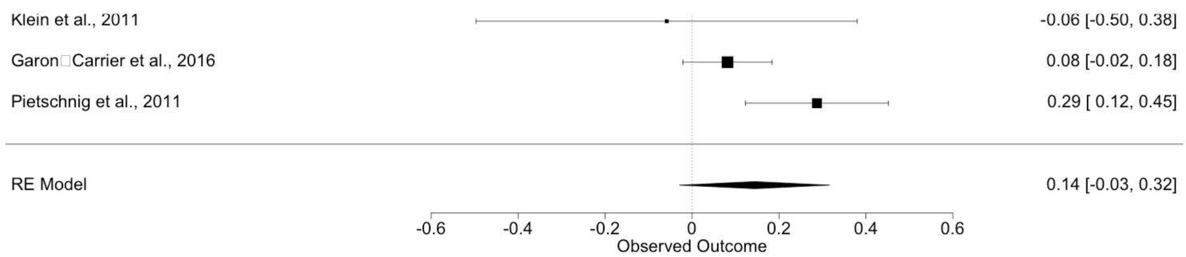


Note. Figure 5 represents the descriptive statistics for this math content.

4.4.3 Basic Numeracy

There was no significant result for basic numeracy tests throughout the studies ($\beta = 0.05$, $SE = 0.13$, 95% CI [-0.20, 0.31]). Forest plot of this content demonstrates the diamond shape intersecting with the vertical line so that means it is not significant. Subsequently, $d = 0.14$ with a confidence interval 95% CI [-0.03, 0.32]. On the other hand, this math content does not include enough studies for comparison and examination.

Figure 6. Forest plot for basic numeracy

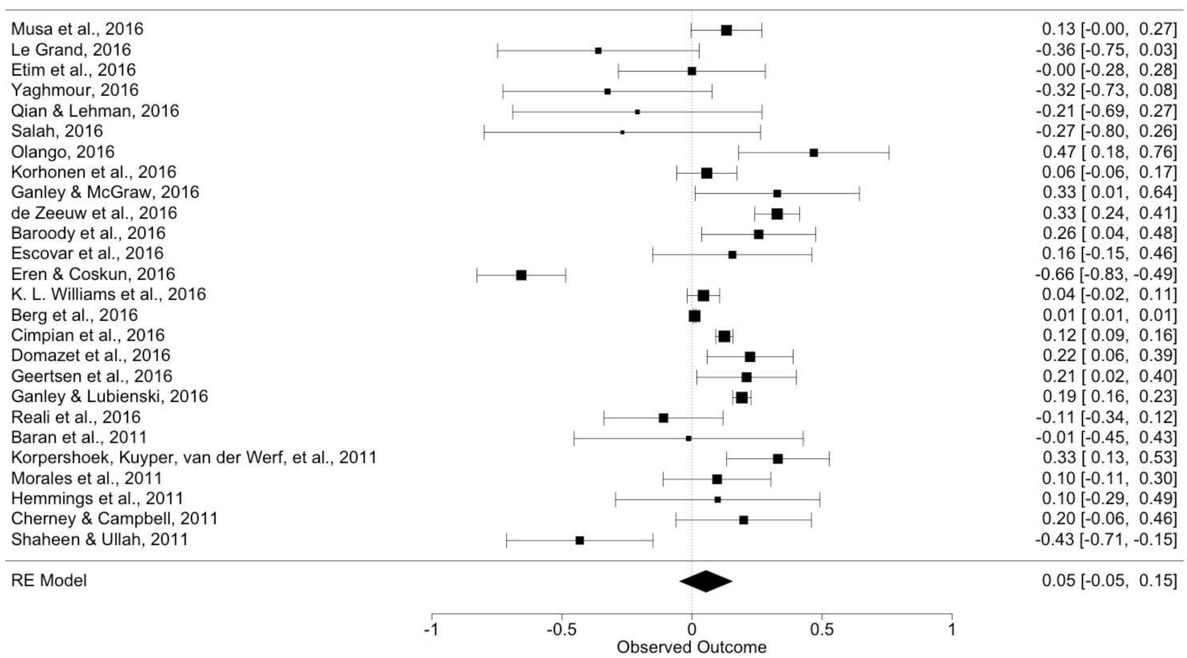


Note. Figure 6 represents the descriptive statistics for this math content.

4.4.4. Broad Mathematics

Similarly, broad mathematics did not show any significant result in the model with moderator (*β* = 0.11, *SE* = 0.12, 95% CI [-0.12, 0.35]). Besides, the diamond shape is on the vertical line indicating no significant result. Upon this, effect size is *d* = 0.05 with a confidence interval of 95% CI [-0.05, 0.15]. Nevertheless, studies were quite dispersed at the right side of the plot, and they are mostly positive meaning that males outperform females in the mathematical assessment.

Figure 7. Forest plot for broad mathematics

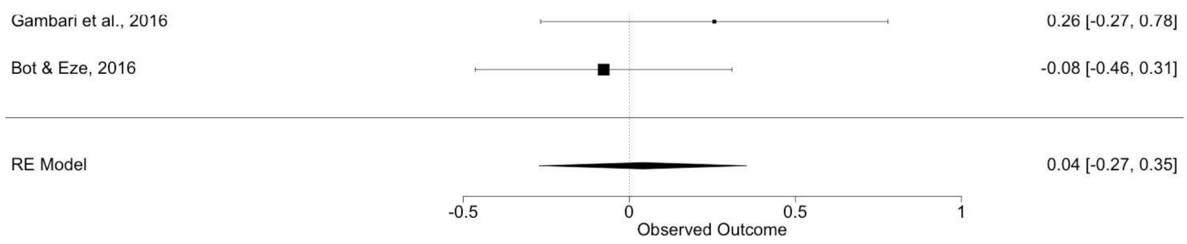


Note. Figure 7 represents the descriptive statistics for this math content.

4.4.5 Geometry

This math content likewise revealed no significant results throughout whole studies in the meta-analysis (*β* < 0.01, *SE* = 0.23, 95% CI [-0.45, 0.46]). Similarly, the forest plot demonstrates similar patterns to other math contents as the diamond shape is located on the vertical line signifying a non-significant result. Additionally, effect size is *d* = 0.04 with a confidence interval 95% CI [-0.27, 035]. Nevertheless, there are only 2 study in this math content section, so this could be the reason for the absence of a significant result.

Figure 8. Forest plot for broad mathematics

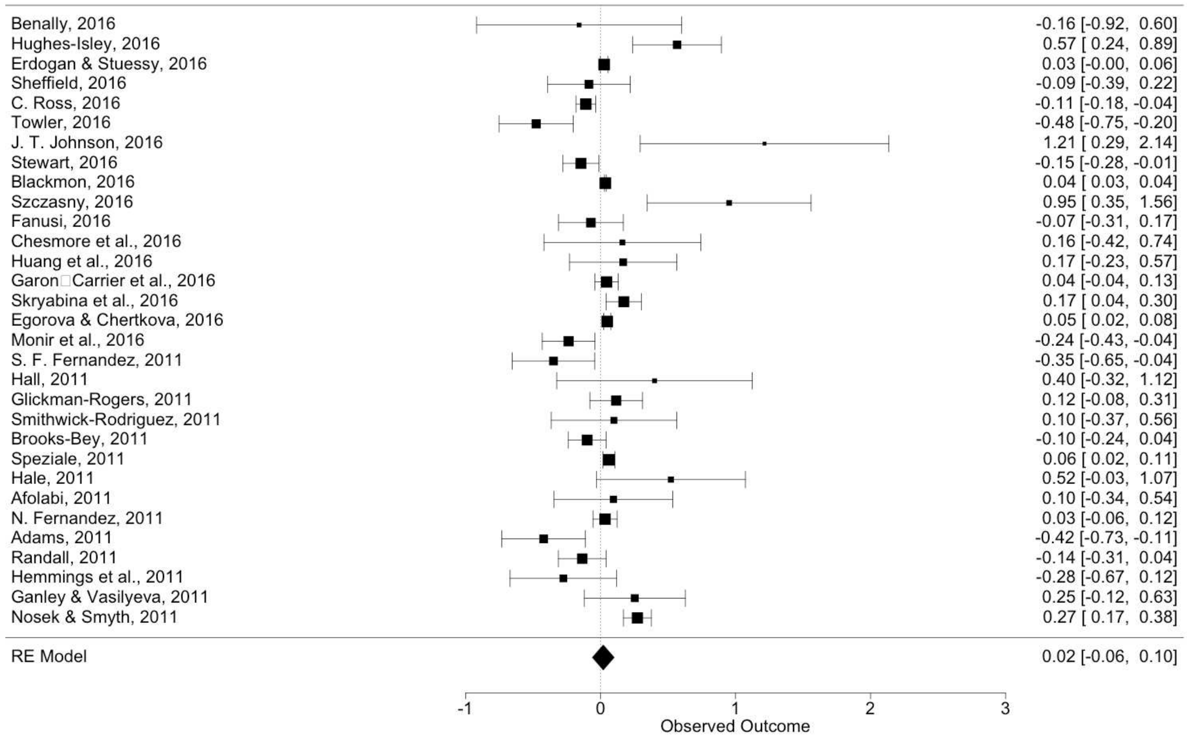


Note. Figure 8 represents the descriptive statistics for this math content.

4.4.6. National Test

Lastly, the national test was examined as a moderator in the study, and similar to other studies, revealed no significant results (β = 0.02, *SE* = 0.12, 95% CI [-0.21, 0.25]). The forest plot displays the results with a diamond shape crossing the vertical line, meaning that the result is not significant. Generally, the studies were not dispersed at both sides of the vertical line but positioned around the vertical line meaning they were not significant. Additionally, the plot shows almost a symmetrical distribution meaning that heteroscedasticity is not present.

Figure 9. Forest plot for national test



Note. Figure 9 represents the descriptive statistics for this math content.

4.5 Overview of results

After model fitting, an analysis was conducted to test whether the two models could be considered statistically different. This analysis has been conducted to examine which model was better through AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) indices. Model with no moderator showed an *AIC* = 126.93 and *BIC* = 150.63, while model with moderator showed an *AIC* = 125.91 and *BIC* = 134.79. Since the second model presents lower *AIC* and *BIC* compared with the first one (*ΔAIC* = 1.02, *ΔBIC* = 15.83), the model with moderator can be considered the best model between the two.

CHAPTER 5

DISCUSSION

The current study aimed to investigate gender differences in mathematics assessments through a meta-analytic study. The study conducted a meta-analysis on published studies from 2010 to 2022, a total of 76 studies extracted as a result based on the established criteria with a total sample of 1,036,286 individuals. This study contributed to the literature review with an emphasis on gender differences in mathematics as a meta-analysis in the current years. Furthermore, particular emphasis was placed on mathematical content to determine if there was an association with gender differences in mathematics.

5.1 Findings

To test whether gender differences exist in math a meta-analysis was carried out. Importantly, publication bias was assessed by mean of an inspection of the funnel plot; the symmetrical distribution of the studies in the funnel plot indicates that publication bias was not present (namely, almost equal number of studies which found a males/females difference in both directions). This suggests that the selection process included a comprehensive sample of published studies, thereby enhancing the reliability and inclusiveness of the study.

The results demonstrated that gender differences emerged in mathematics assessments throughout the studies in 2011 and 2016. The present work found, although negligible, a gender difference in favour of males ($d = 0.08$). These results are in line with previous meta-analyses, specifically, Lindberg and colleagues (2010), found a small gender difference of $d = 0.05$ in a meta-analysis conducted on a systematic review of the literature, and a $d = 0.07$ on large U.S. datasets. Likewise, Else-Quest and colleagues (2010), by analyzing PISA and TIMSS data, found gender differences near to 0 in multiple countries. Altogether, these findings suggest that overall gender gap in math might not exist anymore, and importantly, this trend seems to be stable over time. However, high levels of heterogeneity were detected, potentially due to the presence of confounding variables. Consequently, math content was added in the model as moderator for further investigation of the gender gap. The moderator analyses revealed near-zero gender differences favouring males in all math contents examined, except arithmetic, where males outperformed females solely in the arithmetic tests (d = 0.33). These results were in line with Shen et al. (2016) which highlights the male superiority in complex arithmetic tasks in United States, and Russia. Nevertheless, Wei et al. (2012) found females to achieve higher scores than males in arithmetic tasks including simple subtraction and complex multiplication while adjusting for word-rhyming task. However, the reason behind results was attributed to the fact that females generally display higher performance in language tasks, as the study indicated, which in turn moderates their performance in arithmetic.

Despite some evidence on gender difference in also other domains of mathematics, this study did not reveal any other than arithmetic. For instance, the results proved that males performed better in spatial and abstraction tasks in geometry more than females, indicating that males performed better in tasks that require complex and critical thinking (Fitriani & Nurfauziah, 2019). Conversely, in another study that is conducted in a Turkish sample, females achieved higher scores in geometry compared to males (Erdogan et al., 2011). However, this meta-analysis study did not reveal any results indicating females/males outperform another. For national tests, Lindberg et al. (2010), performed another study on large national databases and found no significant gender gap in mathematics assessments. Similarly, Thomas et al. (2022) conducted another national-test study in Australia and found that males achieved higher scores in numeracy tests compared to females. Conversely, Begum et al. (2021) showed no significant difference of gender for numeracy tests. Additionally, a TIMMS study that is conducted in 2003 showed no significant gender gap in mathematics assessment similarly to our results.

The reason behind general results might be related with the number of studies involves related math content such as advanced mathematics, arithmetic, geometry. As there were limited studies investigating gender differences in some math contents especially for basic numeracy and broad mathematics, the study did not reveal any effect in terms of gender. Even though results might indicate that the gender gap starts to decrease currently it is possible that the sample studies was not sufficient to indicate an effect.

5.2. Limitations and future recommendations

Despite results being useful in the current literature there are some limitations in terms of gender differences in mathematics. Although, the study encompassed a very large sample size, it does not take into consideration cross-cultural factors that might influence gender differences. Else-Quest et al., (2010), revealed a study with TIMSS and PISA databases from 69 countries involving 493,495 students. There are several results regarding gender differences and its associations such as gender equality based on every nation. The study demonstrated that in countries that are more gender-egalitarian, where women' condition and well-being were higher, the gender difference in mathematics was not present as well. Although this meta-analysis study extracted several studies from several countries, it did not take into consideration the level of gender-egalitarian in a country and how results varied accordingly. It could have been another moderator that associate with the results.

Another limitation was that mathematics learning could have been impacted by several aspects such as learning environment, teaching styles, learning environment, peer relationships, parent's expectations, beliefs and stereotypes, parents' education level, teachers' expectations, beliefs and stereotypes, gender stereotypes, cultural environment, role models, problem type, difficulty level of the test and several other factors that needs to be further researched. Since, this study considered only one moderator, it may lack some several other moderators that could be associated with results.

Another topic that should be taken into consideration is aspects such as self-confidence, self-efficacy, self-concept, and math anxiety. Specifically, Else-Quest et al. (2010) revealed that, males display higher levels of positive attitude and positive affect to the mathematics due to some stereotypes against women being capable to perform well. In relation with this, males showed higher self-confidence in mathematics than females and thus mathematics performance can be impacted in favour of males. In addition, females displayed higher anxiety scores when performing mathematics compared to males which may affect their performances. In general, while males' motivation and beliefs, self-report and self-efficacy levels were higher than females. As a result, mathematics performances between females and males are impacted by several aspects and culture. Thus, further studies should investigate this topic by considering some aspects such as culture, stereotypes, anxiety, self-beliefs. Since this study revealed gender differences in arithmetic tests further studies should give emphasis on these aspects when performing in a test of arithmetic.

5.3. Conclusion

There are some notable differences across genders in mathematical assessment in the previous literature. There have been several studies indicating that gender differences are present among mathematics assessment as stated, whereas for some there was no difference. This study revealed gender differences in mathematics assessment with a male advantage, especially in the arithmetic tests. Considering the results, there would be still some discrepancies across genders due to some cognitive and emotive factors that should be further investigated.

REFERENCES

*Adams, C. L. (2011). *The effects of a remedial math intervention on standardized test scores in Georgia middle schools* [Doctoral dissertation, Liberty University] https://core.ac.uk/download/pdf/58825147.pdf

Adnan, M., Lee Abdullah, M.F.N., Puteh, M., Ahmad, C.N.C., & Maat, S.M. (2014). The learning environment and mathematics achievement of students of high performance schools (HPS). *Jurnal Pendidikan Matematik, 2*(1), 1-15.

*Afolabi, K. A. (2011). *Relationship of self-efficacy beliefs of urban public-school students to performance on a high-stakes mathematics test*. [Doctoral dissertation, University of Massachusetts]. Open Access Dissertations. 4348. https://doi.org/10.7275/1667317

Alkharusi, H. (2016). Predicting students' academic achievement: Contributions of perceptions of classroom assessment tasks and motivated learning strategies. *Journal of Research in Educational Psychology, 14*(3), 515-533. http://dx.doi.org/10.14204/ejrep.40.15177

Ayalon, H. & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research, 42,* 432-445. http://dx.doi.org/10.1016/j.ssresearch.2012.10.001

Bacchus, B. (2015). An investigation of the effect of single-gendered instruction on achievement in math and english in an urban high school. *Electronic Theses and Dissertations. 955.* https://egrove.olemiss.edu/etd/955

*Baran, G., Erdogan, S., & Çakmak, A. (2011). A study on the relationship between six-year-old children's creativity and mathematical ability. *International Education Studies*, *4*(1), 105–111. https://doi.org/10.5539/ies.v4n1p105

*Baroody, A. E., Rimm-Kaufman, S. E., Larsen, R. A., & Curby, T. W. (2016). A multi-method approach for describing the contributions of student engagement on fifth grade students' social competence and achievement in mathematics. *Learning & Individual Differences*, *48*, 54–60. https://doi.org/10.1016/j.lindif.2016.02.012

Begum, S., Flowers, N., Tan, K., Carpenter, D.M.H., & Moser, K. (2021). Promoting literacy and numeracy among middle school students: Exploring the mediating role of self-efficacy and gender differences. *International Journal of Educational Research, 106,* 101-722.

*Benally, J. (2015). *Improving Arizona English language learners' mathematics achievement using curriculum-based measures*. [Doctoral Dissertation, Arizona State University]. https://keep.lib.asu.edu/system/files/c7/124470/Benally_asu_0010E_14319.pdf

*Berg, L., Bäck, K., Vinnerljung, B., & Hjern, A. (2016). Parental alcohol-related disorders and school performance in 16-year-olds-a Swedish national cohort study. *Addiction*, *111*(10), 1795–1803. https://doi.org/10.1111/add.13454

*Blackmon, O. M. (2016). *Underrepresented minority students in urban school districts: A study of technology use and student academic performance in math grades four and eight*.

Boaler, J., Altendorff, L., & Kent, G. (2011). Mathematics and science inequalities in the United Kingdom: when elitism, sexism, and culture collide. *Oxford Review of Education, 37*(4), 457-484. http://dx.doi.org/10.1080/03054985.2011.595551

*Bot, T. D., & Eze, J. E. (2016). Comparative effects of concept mapping and cooperative learning strategies on senior secondary school students' achievement in Mathematics-Trigonometry in Kano State, Nigeria. *European Journal of Science and Mathematics Education*, *4*(1), 56–66. https://doi.org/10.30935/scimath/9453

*Brooks-Bey, M. R. (2011). *The impact of the length of preschool attendance on the academic achievement and retention of third and fourth grade students* [Doctoral Dissertation, TUI University]. ProQuest LLC. https://eric.ed.gov/?id=ED534125

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence, 36,* 403-421. https://doi.org/10.1016/j.intell.2007.11.002

Carr, M., Taasoobshirazi, G., Stroud, R., & Royer, J.M. (2011). Combined fluency and cognitive strategies instruction improves mathematics achievement in early elementary school. *Comtemporary Educational Psychology, 36,* 323-333. https://doi.org/10.1016/j.cedpsych.2011.04.002

*Cherney, I. D., & Campbell, K. L. (2011). A league of their own: Do single-sex schools increase girls' participation in the physical sciences? *Sex Roles:* A Journal of Research, 65(9), 712–724. https://doi.org/10.1007/s11199-011-0013-6

*Chesmore, A., Winston, W., & Brady, S. (2016). Academic behavior and performance among African American youth: Associations with resources for resilience. *Urban Review*, *48*(1), 1–14. https://doi.org/10.1007/s11256-015-0342-9

*Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *Grantee Submission*, *2*(4), 1–19. https://doi.org/10.1177/2332858416673617

Cruz-Rodrigues, C., Barbosa, T., Toledo-Piza, C.M.J., Miranda, M.C., & Bueno, O.F.A. (2014). Neuropsychological characteristics of dyslexic children. *Psicologia: Reflexão e Crítica, 27*(3), 539–546. https://doi.org/10.1590/1678-7153.201427315

*Cucina, J. M., Peyton, S. T., Su, C., & Byle, K. A. (2016). Role of mental abilities and mental tests in explaining high-school grades. *INTELLIGENCE*, *54*, 90–104. https://doi.org/10.1016/j.intell.2015.11.007

*de Zeeuw, E. L., van Beijsterveldt, C. E. M., Glasner, T. J., de Geus, E. J. C., & Boomsma, D. I. (2016). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. *Learning & Individual Differences*, *47*, 156–166. https://doi.org/10.1016/j.lindif.2016.01.009

*Domazet, S. L., Tarp, J., Huang, T., Gejl, A. K., Andersen, L. B., Froberg, K., & Bugge, A. (2016). Associations of physical activity, sports participation and active commuting on mathematic performance and inhibitory control in adolescents. *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0146319

*Edwards-Omolewa, N. D. (2011). Gender differences in elementary school children's strategy use and strategy preferences on multidigit addition and subtraction story problems. *Investigations in Mathematics Learning*, *3*(3), 1–25. https://doi.org/10.1080/24727466.2011.11790304

*Egorova, M. S., & Chertkova, Y. D. (2016). Sex differences in mathematical achievement: Grades, national test, and self-confidence. *Psychology in Russia: State of the Art*, *9*(3), 4–23. https://doi.org/10.11621/pir.2016.0301

Elbaum, B. (2007). Effects of an oral testing accodomation on the mathematics performance of secondary students with and without learning disabilities. *The Journal of Special Education, 40*(4), 218-229. https://doi.org/10.1177/00224669070400040301

Else-Quest, N.M., Hyde, S.H., & Linn, M.C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. https://doi.org/10.1037/a0018053

Erdogan, F. (2020). Prospective middle school mathematics teachers' problem posing abilities in context of Van Hiele levels of geometric thinking. *International Online Journal of Educational Services, 12*(2), 132-152. https://doi.org/10.15345/iojes.2020.02.009

Erdogan, A., Baloglu, M., & Kesici, Ş. (2011). Gender differences in geometry and mathematics achievement and self-efficacy beliefs in geometry. *Eurasian Journal of Educational Research, 43,* 188-205. https://doi.org/10.1017/CBO9780511614446.015

*Erdogan, N., & Stuessy, C. (2016). Examining the role of inclusive STEM schools in the college and career readiness of students in the United States: A multi-group analysis on the outcome of student achievement. *Educational Sciences: Theory and Practice*, *15*(6), 1517–1529. https://doi.org/10.12738/estp.2016.1.0072

*Eren, A., & Coskun, H. (2016). Students' level of boredom, boredom coping strategies, epistemic curiosity, and graded performance. *Journal of Educational Research*, *109*(6), 574–588. https://doi.org/10.1080/00220671.2014.999364

*Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., & Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Scientific Reports*, *6*, 23011. https://doi.org/10.1038/srep23011

*Etim, J. S., Etim, A. S., Heilman, G., Mathiyalakan, S., & Ntukidem, E. (2016). Gender and achievement in English language arts, science and mathematics in secondary schools in Nigeria. *European Journal of Science and Mathematics Education*, *4*(2), 186–195. https://doi.org/10.30935/scimath/9464

*Ezeugwu, J. J. O., Onuorah, J. C., Asogwa, U. D., & Ukoha, I. P. (2016). Effect of mathematics game-based instructional techniques on students'achievements and interest in algebra at basic education level. *Global Journal of Pure and Applied Mathematics*, *12*(4), 3727–3744.

Fahle, E. (2016). *Patterns of change in U.S. gender achievement gaps during elementary and middle school* [Doctoral dissertation, Stanford University]. ERIC. https://eric.ed.gov/?id=ED567028

*Fanusi, A. D. (2016). *The effect of ALEKS math support on standardized math test scores in middle school*. [Doctoral dissertation, Liberty University]. ProQuest Dissertations.

*Fernandez, N. (2011). *Early childhood education: The sustainability of the benefits of preschool participation in Abbott districts*. [Doctoral dissertation, Seton Hall University]. ProQuest LLC. 2743. https://scholarship.shu.edu/dissertations/379

*Fernandez, S. F. (2011). *Brain Hemisphericity and Mathematics Achievement of High School Students*. [Master's Thesis, West Visayas State University]. WVSU Institutional

Repository and Electronic Dissertations and Theses PLUS. https://repository.wvsu.edu.ph/handle/20.500.14353/147

Fischer, U., , K., Bientzle, M., Cress, U., & Christoph Nuerk, H. (2011). Sensori-motor spatial training of number magnitude representation. *Psychon Bull Rev, 18,* 177-183. https://doi.org/10.3758/s13423-010-0031-3

Fitriani, N., & Nurfauziah, P. (2019). Gender and mathematical abstraction on geometry. *Journal of Physics: Conference Series, 1315(1), 012052*. doi:10.1088/1742-6596/1315/1/012052

*Gambari, A. I., Shittu, A. T., Daramola, F. O., & James, M. (2016). Effects of video-based cooperative, competitive and individualized instructional strategies on the performance of senior secondary schools students in Geometry. *Malaysian Online Journal of Educational Sciences*, *4*(4), 31–47.

*Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning & Individual Differences*, *47*, 182–193. https://doi.org/10.1016/j.lindif.2016.01.002

*Ganley, C. M., & McGraw, A. L. (2016). The development and validation of a revised version of the Math anxiety scale for young children. *Front Psychol*, *7*, 1181. https://doi.org/10.3389/fpsyg.2016.01181

*Ganley, C. M., & Vasilyeva, M. (2011). Sex differences in the relation between math performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology*, *32*(4), 235–242. https://doi.org/10.1016/j.appdev.2011.04.001

*Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J.-P., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *CHILD DEVELOPMENT*, *87*(1), 165–175. https://doi.org/10.1111/cdev.1245

Geary, D.C. (2011). Cognitive predictors of achievement growth in mathematics: A five-year longitudinal study. *Dev Psychol, 47*(6), 1539-1552. https://doi.org/10.1037/a0025510

*Geertsen, S. S., Thomas, R., Larsen, M. N., Dahn, I. M., Andersen, J. N., Krause-Jensen, M., Korup, V., Nielsen, C. M., Wienecke, J., Ritz, C., Krustrup, P., & Lundbye-Jensen, J. (2016). Motor skills and exercise capacity are associated with objective measures of cognitive functions and academic performance in preadolescent children. *PLOS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0161960

Geist, E.A. & King M. (2008). Different, not better: Gender differences in mathematics learning and achievement. *Journal of Instructional Psychology, 35(1),* 43-52.

*Glickman-Rogers, A. (2011). *How do social and emotional competencies influences the grades that male students receive?* 1537.

Goldin, G. (2008). Perspectives on Representation in Mathematical Learning and Problem Solving. In B. English, L. D. & Kirshner, D. (Eds.), *Handbook of International Research in Mathematics Education* (pp 176-202). Routhledge International Handbooks.

*Hale, J. V. (2011). *An examination of reading and mathematic achievement among second grade students who have received instruction from either teachers who have been trained in choice theory/reality therapy methods or teachers who have not been trained.*

[Doctoral dissertation, Duquesne University]. ProQuest LLC. https://eric.ed.gov/?id=ED533711

*Hall, A. R. (2011). *College readiness: The evaluation of students participating in the historically black college and university program in Pre-Calculus and the Calculus sequence*. [Doctoral dissertation, Southern University and Agricultural and Mechanical College]. ProQuest LLC. https://eric.ed.gov/?id=ED536058

Hanson, M.J., Miller, A.D., Diamond, K., Odom, S., Lieber, J., Butera, G., Horn, E., Palmer, S., & Fleming, K. (2011). Neighborhood community risk influences on preschool children's development and school readiness. *Infants & Young Children, 24*(1), 87-100. https://doi.org/10.1097/IYC.0b013e3182008dd0

*Hemmings, B., Grootenboer, P., & Kay, R. (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *INTERNATIONAL JOURNAL OF SCIENCE AND MATHEMATICS EDUCATION*, *9*(3), 691–705. https://doi.org/10.1007/s10763-010-9224-5

*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 & Individual Differences*, *48*, 45–53. https://doi.org/10.1016/j.lindif.2016.02.006

*Hopf, F. C. (2011). *The impact of a short-term review treatment program on student success in a college algebra course*. 3677. [Doctoral dissertations, University of South Florida]. Digital Commons. https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=4348&context=etd

*Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning & Individual Differences*, *47*, 258–265. https://doi.org/10.1016/j.lindif.2016.01.012

*Hughes-Isley, S. (2016). *Differences between male and female community college students in achievement and attitude on college remedial mathematics*. [Doctoral dissertation, Colorado State University]. https://mountainscholar.org/items/9edc367a-d2fa-471c-820d-ee6fc2d07c16

*Jansen, B. R. J., Hofman, A. D., Savi, A., Visser, I., & van der Maas, H. L. J. (2016). Self-adapting the success rate when practicing math. *Learning & Individual Differences*, *51*, 1–10. https://doi.org/10.1016/j.lindif.2016.08.027

*Jõgi, A.-L., & Kikas, E. (2016). Calculation and word problem-solving skills in primary grades - Impact of cognitive abilities and longitudinal interrelations with task-persistent behaviour. *BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY*, *86*(2), 165–181. https://doi.org/10.1111/bjep.12096

*Johnson, J. T. (2016). *A study of the impact of retention on student achievement in three rural Missouri school districts*. [Doctoral Dissertation, Lindenwood University]. Theses & Dissertations at Digital Commons@Lindenwood University. https://digitalcommons.lindenwood.edu/dissertations/374

*Klein, P. S., Adi-Japha, E., & Hakak-Benizri, S. (2010). Mathematical thinking of kindergarten boys and girls: Similar achievement, different contributing processes. *EDUCATIONAL STUDIES IN MATHEMATICS*, *73*(3), 233–246. https://doi.org/10.1007/s10649-009-9216-y

*Konan, P. N., Chatard, A., Selimbegović, L., Mugny, G., & Moraru, A. (2011). Deflecting stereotype threat through downward comparison: When comparison with immigrants boosts the performance of stigmatized native students. *Social Justice Research*, *24*(2), 191–205. https://doi.org/10.1007/s11211-011-0134-7

Konstantopoulos, S. & Chung, V. (2011). Teacher effects on minority and disadvantaged students' grade 4 achievement. *The Journal of Educational Research, 104*(2), 73-86. http://dx.doi.org/10.1080/00220670903567349

*Korhonen, J., Tapola, A., Linnanmäki, K., & Aunio, P. (2016). Gendered pathways to educational aspirations: The role of academic self-concept, school burnout, achievement and interest in mathematics and reading. *LEARNING AND INSTRUCTION*, *46*, 21–33. https://doi.org/10.1016/j.learninstruc.2016.08.006

*Korpershoek, H., Kuyper, H., van der Werf, G., & Bosker, R. (2011). Who succeeds in advanced mathematics and science courses? *BRITISH EDUCATIONAL RESEARCH JOURNAL*, *37*(3), 357–380. https://doi.org/10.1080/01411921003671755

Lambert, R. & Tan, P. (2016, November). Disability and mathematics: Theorizing the research divide between special education and mathematics. In Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.), *38th North American Chapter of the International Group for the Psychology of Mathematics Education.* Tucson, AZ: University of Arizona. *https://eric.ed.gov/?id=ED583739*

*Le Grand, A. J. (2016). *Gender differences in children's academic self-concept and achievement in math and reading*. [Doctoral dissertation, Northern Arizona University]. ProQuest Dissertations.

https://www.proquest.com/openview/820ea00a12e6d86132bceda9b6e8dbd3/1?pq-origsite=gscholar&cbl=18750

Lindberg, S.M., Hyde, J.S., & Petersen, J.L. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychology Bulletin, 136*(6), 1123-1135. https://doi.org/10.1037/a0021276

Malik, M.I., Farooq, R.A., & Tabassum, R. (2016). The comparative study of achievement of male and female mathematics students of higher secondary schools and colleges at intermediate level in Punjab. *Bulletin of Education and Research, 38*(2), 219-227.

Mann, A. & DiPrete, T.A. (2016). The consequences of the national math and science performance environment for gender differences in STEM aspiration. *Sociological Science, 3,* 568-603. https://doi.org/10.15195/v3.a25

Mc-Millian, M.M., Frierson, H.T., & Campbell, F.A. (2011). Do gender differences exist in the identification of African American elementary school aged children? *J Black Psychol, 37*(1), 78-98. https://doi.org/10.1177/0095798410366709v

*Mendiburo, M. & Hasselbring, T. (2011). Technology's impact on fraction learning: An experimental comparison of virtual and physical manipulatives. *Society for Research on Educational Effectiveness*. https://eric.ed.gov/?id=ED517987

*Mittal, S., Verma, P., Jain, N., Singh, A. K., Khatter, S., & Juyal, A. (2011). Gender based variation in cognitive functions in adolescent subjects. *Indian Journal of Physiology and Pharmacology*, *55*(5), 294–295. https://doi.org/10.5214/ans.0972.7531.190406

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.C (2011). Early place-value understanding as a precursor for later arithmetic performance-A longitudinal study on numerical development. *Research in Developmental Disabilities, 32,* 1837-1851.

*Monir, Z. M., Salah El-Din, E. M., El-Alameey, I. R., Yamamah, G. A., Megahed, H. S., Salem, S. M., & Ibrahim, T. S. (2016). Academic achievement and psychosocial profile of Egyptian primary school children in South Sinai. *Open Access Macedonian Journal of Medical Sciences*, *4*(4), 624. https://doi.org/https://doi.org/10.3889/oamjms.2016.111

*Morales, J., González, L.-M., Guerra, M., Virgili, C., & Unnithan, V. (2011). Physical activity, perceptual-motor performance, and academic learning in 9- to-16-years-old school children. *International Journal of Sport Psychology*, *42*(4), 401–415.

*Musa, A. K. J., Dauda, B., & Umar, M. A. (2016). Gender differences in achievement goals and performances in English language and mathematics of senior secondary school students in Borno State, Nigeria. *Journal of Education and Practice*, *7*(27), 165–175. https://eric.ed.gov/?id=EJ1115862

Navarro, J.I., Aguilar, M., Alcalde, C., Ruiz, G., Marchena, E., & Menacho, I. (2011). Inhibitory process, working memory, phonological awareness, naming speed, and early arithmetic achievement. *The Spanish Journal of Psychology, 14*(2), 580-588. http://dx.doi.org/10.5209/rev_SJOP.2011.v14.n2.6aş

Niederle, M. & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives, 24*(2), 129-144.

Nielsen, H.S. & Rangvid, B.S. (2012). The impact of parents' years since migration on children's academic achievement. *Journal of Migration, 1*(6).

*Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125–1156. https://doi.org/10.3102/0002831211410683

*Okpube, N. M., & Anugwo, M. N. (2016). Card games and Algebra tic tacmatics on achievement of junior secondary II students in Algebraic expressions. *International Journal of Evaluation and Research in Education*, *5*(2), 93–100. https://doi.org/10.11591/ijere.v5i2.4527

*Olango, M. (2016). Mathematics anxiety factors as predictors of mathematics self-Efficacy and achievement among freshmen science and engineering students. *African Educational Research Journal*, *4*(3), 109–123. https://eric.ed.gov/?id=EJ1216175

Oswald, T.M., Beck, J.S., Losif, A.M., McCauley, J.B., Gilhooly, L.J., Matter, J.C., & Solomon, M. (2016). Clinical and cognitive characteristics associated with mathematics problem-solving in adolescents with autism spectrum disorder. *Autism Res, 9*(4), 480-490. https://doi.org/10.1002/aur.1524

Passolunghi, M.C. (2011). Cognitive and emotional factors in children with mathematical learning disabilities. *International Journal of Disability, Development and Education, 58*(1), 61-73. https://10.1080/1034912X.2011.547351

Pegg, J. E. (2003). Assessment in mathematics: A developmental approach. In B. J. M.Royer (Ed.), *Mathematical Cognition* (pp 227-259). Information Age Pubslihing.

Penner, A.M. & Paret, M. (2007). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research, 37,* 239-253.

*Pietschnig, J., Voracek, M., & Formann, A. K. (2011). Female Flynn effects: No sex differences in generational IQ gains. *Personality And Individual Differences*, *50*(5), 759–762. https://doi.org/10.1016/j.paid.2010.12.019

Plaisted, K., Bell, S., & Mackintosh, N.J. (2011). The role of mathematical skill in sex differences on Raven's Matrices. *Personality and Individual differences, 51,* 562-565. https://doi.org/10.1016/j.paid.2011.05.005

*Qian, Y., & Lehman, J. D. (2016). Correlates of success in introductory programming: A Study with middle school students. *Journal of Education and Learning*, *5*(2), 73–83. https://doi.org/10.5539/jel.v5n2p73

*Randall, P. P. (2011). *Preschool education in Virginia and the resulting academic effects for third- and fifth-grade at-risk students*. [Doctoral dissertation, Liberty University]. ProQuest LLC. 4279. https://eric.ed.gov/?id=ED523365

*Reali, F., Jiménez-Leal, W., Maldonado-Carreño, C., Devine, A., & Szücs, D. (2016). Examining the link between math anxiety and math performance in Colombian students. *Revista Colombiana de Psicología*, *25*(2), 369–379. https://doi.org/10.15446/rcp.v25n2.54532

Reardon, S., Fahle, E., Kalogrides, D., Podolsky, A., & Zarate, R. (2016). Test format and the variation of gender achievement gaps within the Unites States. *Society for Research on Educational Effectiveness.* https://eric.ed.gov/?id=ED567042

Reilly, D., Neumann, D.L., & Andrews, G. (2019). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education, 49*(4), 25-50. https://doi.org/10.1007/s11165-017-9630-6

Robinson, J.P. & Lubienski, S.T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48*(2), 268-302.

Ross, P. & Randolph, J. (2014). Differences between students with and without ADHD on task vigiliance under conditions of distraction. *Journal of Educational Research and Practice, 4*(1), 1- https://doi.org/10. *10.5590/JERAP.2014.04.1.01*

*Ross, C. (2016). *Title I versus Non-Title I Students and Standardized Test Scores in a School District in Georgia*. [Doctoral dissertation, Capella University]. ProQuest LLC. https://eric.ed.gov/?id=ED570277

*Salah, R. O. (2016). Effectiveness of using computerized educational packages in teaching math curriculum on the learning of student in eighth grade at Ma'an City. *Journal of Education and Practice*, *7*(12), 60–66.

*Saleh, S., & Rahman, M. A. A. (2016). A study of students' achievement in algebra: Considering the effect of gender and types of schools. *European Journal of STEM Education*, *1*(1), 19–26. https://doi.org/10.20897/lectito.201603

Sandoval-Hernandez, A. & Bialowolski, P. (2016). Factors and conditions promoting academic resilience: a TIMSS-based analysis of five Asian education systems. *Asia Pasific Educ. Rev, 17*, 511-520. https://doi.org/10.1007/s12564-016-9447-4

Sari, D.N.O. (2020). Gender differences in junior high school students' mathematical connection in geometry. *Ahmad Dahlan International Conference on Mathematics and Mathematics Education.* IOP Publishing. J. Phys.: Conf. Ser. 1613 012069. https://doi.org/10.1088/1742-6596/1613/1/012069

*Shaheen, A., & Ullah, S. (2011). Cognitive function in children. Gander specific difference in general public primary school children in Pakistan. *European Psychiatry*, *26*. https://doi.org/10.1016/S0924-9338%2811%2972058-2

*Sheffield, L. E. (2016). *The impact of positive behavior intervention support on male and female students' academic performance*. [Doctoral dissertation, South Carolina State University]. ProQuest LLC. https://eric.ed.gov/?id=ED589185

Shen, C., Vasilyeva, M., & Laski, E.V. (2016). Here, but not there: Cross-national variability of gender effects in arithmetic. *Journal of Experimentel Child Psychology, 146,* 50-65.

*Skryabina, E., Taylor, G., & Stallard, P. (2016). Effect of a universal anxiety prevention programme (FRIENDS) on children's academic performance: results from a randomised controlled trial. *Journal of Child Psychology & Psychiatry*, *57*(11), 1297–1307. https://doi.org/10.1111/jcpp.12593

*Smithwick-Rodriguez, T. (2011). *A study of college readiness programs among high school students in South Texas* [Doctoral dissertation, Texas A&M University]. ProQuest LLC. https://eric.ed.gov/?id=ED535796

Smith, S.J. (2011). *The comparison of single-sex classrooms on the achievement of reading and math for grade five students.* [Doctoral dissertation, Trident University]. ProQuest LLC. https://www.proquest.com/docview/880204048/

*Speziale, K. M. (2011). *The impact of birth month, within grade levels, on student performance on the Pennsylvania system of school assessment (PSSA) mathematics assessment*. [Doctoral dissertation, Wilkes University]. ProQuest LLC. https://eric.ed.gov/?q=Sex&ff1=eduGrade+5&ff2=eduGrade+11&id=ED534200

*Stewart, R. L. (2016). *The relationships between the Teacher-Insight(TM) score and student performance as measured by Student TAKS academic change scores* [Doctoral dissertation, University of North Texas]. UNT Digital Library. https://digital.library.unt.edu/ark:/67531/metadc700102/m2/1/high_res_d/dissertation.pdf

Sutton, A.M. (2009). *The impact of single-gender classrooms on student achievement in seventh grade math classess.* [Doctoral dissertation, South Carolina University]. ProQuest LLC. https://eric.ed.gov/?id=ED527041

*Swearingen, D. K. (2011). *Effect of digital game based learning on ninth grade students' mathematics achievement*. [Doctoral Dissertation, The University of Oklahoma]. ProQuest LLC. https://eric.ed.gov/?id=ED528934

*Szczasny, A. L. (2016). *A Study of the Effect of the "Brain Gym" Intervention on the Math and Reading Achievement Scores of Fourth Grade Students*. [Doctoral dissertation, University of St. Francis]. ProQuest LLC. https://eric.ed.gov/?id=ED589185

*Tarazi, N. (2016). *The influence of the inverted classroom on student achievement and motivation for learning in secondary mathematics in the United Arab Emirates: A quasi-experimental study*. [Doctoral dissertation, Northcentral University]. ProQuest LLC. https://eric.ed.gov/?id=ED589238

Thomas, D.P., Hopwood, B., Hatisaru, V., & Hicks, D. (2022). Gender differences in reading and numeracy achievement across the school years. *The Australian Education Researcher,* 51(1), 41-46. https://doi.org/10.1007/s13384-022-00583-8

Tobia, V., Bonifacci, P., & Marzocchi, G.M. (2014). Concurrent and longitudinal predictors of calculation skills in preschoolers. *European Journal of Psychology of Education, 31*(2), 155-174. https://doi.org/10.1007/s10212-015-0260-y

*Towler, S. M. (2016). *A comparison of academic achievement of urban students in 21 st cclc after school programs to students not enrolled*. [Doctoral Dissertation, Capella University]. ProQuest Dissertations. https://doi.org/10.30827/Digibug.54021

Uttal, D.H. (1997). Beliefs about genetic influences on mathematics achievement: a cross-cultural comparison. *Genetica, 99*(2-3), 165-72. https://doi.org/10.1007/BF02259520

Vergnaud, G. (1998). A comprehensive theory of representation for mathematics education. *Journal of Mathematical Behavior, 17*(2), 167-181. https://doi.org/10.1016/S0364-0213(99)80057-3

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q. & Zhou, X. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science, 23*(3), 320-330.

*Williams, K. L., Burt, B. A., & Hilton, A. A. (2016). Math achievement: A role strain and adaptation approach. *JOURNAL FOR MULTICULTURAL EDUCATION*, *10*(3), 368–383. https://doi.org/10.1108/JME-01-2016-0005

*Yaghmour, K. S. (2016). Effectiveness of blended teaching strategy on the achievement of third grade students in mathematics. *Journal of Education and Practice*, *7*(5), 65–73. https://eric.ed.gov/?id=EJ1092394