AYTEN SENGÜ TOSUN

2022

META-ANALYSIS ON THE EFFECT OF ENRICHMENT PROGRAMS ON THE ACADEMIC ACHIEVEMENT OF GIFTED AND TALENTED STUDENTS

AYTEN SENGÜ TOSUN

A THESIS SUBMITTED

FOR

THE DEGREE OF MASTER OF ARTS

IN

CURRICULUM AND INSTRUCTION

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

AUGUST 2022

To the three E's in our family,
Thanks for supporting me in this long, fun, and never seems to be ending journey.

Meta-analysis on the Effect of Enrichment Programs on the Academic Achievement of Gifted and Talented Students

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Ayten Sengü Tosun

In Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in

Curriculum and Instruction

Ankara

August 2022

İHSAN DOĞRAMACI BILKENT UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

Meta-analysis on the Effect of Enrichment Programs on the Academic Achievement of Gifted and Talented Students

Ayten Sengü Tosun

July 2022

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

Assoc. Prof. Dr. Erdat Çataloğlu (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

Assoc. Prof. Dr. Jennie Faber Lane (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

Prof. Dr. Salih Ateş, Gazi University (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Orhan Arıkan (Director)

ABSTRACT

META-ANALYSIS ON THE EFFECT OF ENRICHMENT PROGRAMS ON THE ACADEMIC ACHIEVEMENT OF GIFTED AND TALENTED STUDENTS

Ayten Sengü Tosun

MA in Curriculum and Instruction

Advisor: Assoc. Prof. Dr. Erdat Çataloğlu

August 2022

Much research has been conducted about gifted education recently. One of the most applied interventions in gifted education is enrichment. Enrichment is the teaching and learning process that increases the depth and complexity of typical school subjects. In this thesis, the effects of enrichment on the academic achievement of gifted and talented students, was examined. The meta-analytic procedure was used to determine the overall effect size. The researcher used Education Resources Information Center database (ERIC), the most comprehensive database on education literature, Bilkent online library, and Google Scholar to detect relevant studies. The studies were filtered to the studies published between 2013 and 2021. After applying the inclusion and exclusion criteria, 22 studies were determined to be included in this meta-analysis. The pooled effect size was estimated under the random-effects model using the R statistical software. The effect size measure used was Hedges' g. The findings showed that enrichment programs had a positive moderate effect (g= 0.68, 95% CI [0.31, 1.05]) on students' academic achievement. To investigate the robustness of the findings, heterogeneity analyses were conducted and publication bias was investigated through the funnel plot. High heterogeneity (I^2=84.6%) of the findings supports the use of a random-effects model. Also, the funnel plot showed an asymmetry which might be the reason for this heterogeneity and/or a possible publication bias found in the literature. In addition, subgroup analyses were performed to check for possible reasons for this high heterogeneity.

Keywords: *Gifted and talented students, talent development, enrichment, academic achievement, meta-analysis*

ÖZET

Zenginleştirilmiş Programların Üstün Zekâlı ve Yetenekli Öğrencilerin Akademik Başarısına Etkisi Üzerine Meta-analiz Çalışması

Ayten Sengü Tosun

Eğitim Programları ve Öğretim Yüksek Lisans Programı

Danışman: Doç. Dr. Erdat Çataloğlu

Ağustos 2022

Son zamanlarda üstün zekâlı ve yetenekli eğitimi alanında birçok akademik çalışma yapılmıştır. Bu alanda en çok uygulanan iki uygulama zenginleştirme ve hızlandırmadır. Zenginleştirme, öğretim ve öğrenim sürecinin derinliğini ve karmaşıklığının arttırılmasıdır. Bu tezde, zenginleştirmenin üstün zekâlı ve yetenekli öğrencilerin akademik başarısına etkileri incelenmektedir. Genel etki büyüklüğünün belirlenmesi için yöntem olarak meta-analizi belirlenmiştir. Konuyla alakalı daha önce yapılan çalışmaların belirlenmesi için ERIC, Bilkent Üniversitesi ve Google Scholar veri tabanları kullanılmıştır. Tarama sırasında çalışmalar 2013 ve 2021 yılları arasında yayınlanan çalışmalarla sınırlandırılmıştır. Çalışmaya dâhil etme ve hâriç tutma kriterlerine göre çalışmalar sınıflandırıldıktan sonra, analiz edilmek üzere 22 çalışma saptanmıştır. Ortalama etki büyüklüğü, rastgele etkiler modeline göre R istatistik program kullanılarak hesaplanmıştır. Hesaplanan etki değeri Hedges' g türündendir. Sonuçlar zenginleştirmenin üstün zekâlı ve yetenekli öğrencilerin akademik başarısına pozitif yönde orta etki büyüklüğünde etki büyüklüğüne (g= 0.68, %95 CI [0.31, 1.05]) sahip olduğunu göstermektedir. Çalışmanın sonuçlarının sağlamlığının incelenmesi için çeşitli heterojenlik testler yapımlı ve yayın yanlılığı huni diyagramı çizilerek incelenmiştir. Heterojenlik test sonuçlarının yüksek çıkması(I^2=%84,6), çalışmada rastgele etkiler modelinin kullanılmasını destekler nitelikte. Ayrıca, huni diyagramının asimetrik olması yüksek heterojenliği ve/veya yayın yanlılığı ihtimalini ortaya koydu. Bunun yanında, yüksek heterojenliğin muhtemel sebeplerini saptamak için alt grup analizleri yapıldı.

Anahtar kelimeler: *Üstün zekâlı ve yetenekli öğrenciler, yetenek geliştirme, zenginleştirme, akademik başarı, meta-analiz*

ACKNOWLEDGMENTS

I would like to take this opportunity to express my appreciation to those who have guided and supported me throughout my master's journey.

First and foremost, I would like to acknowledge and give my warmest thanks to my supervisor, Assoc. Prof. Dr. Erdat Çataloğlu for encouraging me to try different approaches throughout my research. His constant support and patience turned this challenging journey into a fun, enjoyable learning opportunity. I am and will be grateful for all of your assistance and support.

In addition to my supervisor, I have been thankful to the committee members, Assoc. Prof. Dr. Jennie Farber Lane and Prof. Dr. Salih Ateş for accepting our offer and taking your time to read my thesis thoroughly. Thank you for your valuable, constructive feedback and questions that developed this thesis.

I should convey my gratitude to my favorite math people, Dr. Anita Nicole Alexander and Elif Koç for teaching me a lot. It was always fun learning from and with you. You were great examples of an inspiring mentor and a supportive cohort.

I want to thank my colleagues and friends Selin Gümüşgül and Damla Sena Ercazip for giving me the motivation to pursue. Also, I wish to thank my dearest friends Betül Döner and Göktuğ Yılbaşı for being there for me and pushing me to finish this thesis.

I would also like to thank Anıl Yıldız, for always encouraging me. Your constant support kept me writing this thesis. Thanks for sharing all the stress with me and believing me at all costs.

Last but not least, I want to thank my parents, Esin and Ersan Tosun, for their love and unwavering support. Thank you for always providing me with options and allowing me to make my own decisions. This thesis is dedicated to both of you and all the times we enjoyed together as a family.

TABLE OF CONTENTS

ABSTRACT iii

ÖZET iv

ACKNOWLEDGMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES x

LIST OF FIGURES xii

CHAPTER 1: INTRODUCTION 1

Introduction 1

Background 1

General Discussion on Modern Schooling 1

Gifted Education in the Past 3

Major Contributors to the Gifted Education Field 3

Effects of Sputnik's Launch on Gifted Education 5

Problem 7

Purpose 12

Research Question 12

Significance 13

Limitations 15

Definition of Key Terms 17

CHAPTER 2: LITERATURE REVIEW 19

Introduction 19

Gifted and Talented Education 19

Enrichment Programs 22

Enrichment Models 23

Enrichment Triad Model 23
The Schoolwide Enrichment Model (SEM) 24
Three Stage Enrichment Model 25
Levels of Service Model 25
Strategies Used in Enrichment Activities 26
Independent Studies 26
Learning Centers 26
Saturday and After School Programs 27
Academic Competitions 27
Academic Achievement 27
Meta-analysis 28
Advantages of Meta-analysis 29
Disadvantage of Meta-analysis 30
Examples of Meta-analyses on Gifted Education 31
Conclusion 34
CHAPTER 3: METHOD 35
Introduction 35
Research Method 36
Meta-analysis 36
Effect Size Measures and Types 37
Determination of Effect Size 38
Meta-analytical Models 41
The Fixed-effect Model 41
The Random-effects Model 42
Conducting Meta-analysis 42

Step 1: Formulating the Research Question 43
Step 2: Literature Search 44
Step 3: Choosing of Effect Size Measure 45
Step 4: Choosing the Analytical Method 45
Step 5: Choosing Software 45
Step 6: Creating the Dataset 46
Step 7: Statistical Computation 46
Step 8: Reporting Results 47
Meta-analysis Conducted 47
Research Question 48
Literature Search 49
Inclusion and Exclusion Criteria 52
Electronic Database Search 52
Effect Size Measure Choice 53
Meta-analytical Model Choice 53
Software Selection 55
Data Extraction 55
Treatment of Multiple Effect Sizes in a Study 55
Analysis 56
CHAPTER 4: RESULTS 58
Introduction 58
Results of the Study 59
Computing the Overall Effect Size 59
Heterogeneity Analysis 62
Investigating Possible Outliers and Influential Cases 63

Brute Force Method 63

Leave-one-out Method 65

Graphic Display of Heterogeneity (GOSH) Plot Analysis 67

Subgroup Analysis and Hypothesis Testing 69

Testing Publication Bias 76

Funnel Plot 77

Duval & Tweedie Trim and Fill Method 78

CHAPTER 5: DISCUSSION 82

Introduction 82

Overview of the Study 82

Discussion of Major Findings 84

Overall Conclusion 90

Recommendations for Future Research 91

REFERENCES 93

Appendix A 107

Appendix B 108

LIST OF TABLES

Table Page

1 Three Stage Model 25

2 The Timetable of Thesis 35

3 Forms of Effect Sizes (ES) Used Commonly in a Meta-Analysis (Continuous Data) 40

4 PICO for This Thesis 49

5 Number of Search Results 51

6 Included Studies 54

7 τ^2 and I^2 Heterogeneity Analyses 62

8 Cochran's Q Heterogeneity Measure 63

9 Pooled Effect Size of Studies after Removing Outliers (Random Effects Model) 64

10 Heterogeneity Analyses after Removing Outliers 64

11 Cochran's Q Heterogeneity Measure after Removing Outliers 65

12 Variables to Check when Exploring the GOSH Plot 68

13 Sample Size of the Subgroups 71

14 Subgroup Analysis by Sample Size 71

15 Subgroup Analysis by Grade Levels 72

16 Subgroup Analysis by the Publication Language 72

17 Subgroup Analysis by the Study Design 73

18 Subgroup Analysis of Journal and Dissertations 74

19 Subgroup Analysis of Studies by the Tests They Use 74

20 Subgroup Analysis of studies by the Subject Area 75

21 Summary Results of Hypothesis Testing 76

22 Pooled Effect Size after Trim and Fill Procedure 79

23 Pooled Effect Size after Trim and Fill Procedure without Outliers 81

24 Investigating Possible Outliers..... 91

LIST OF FIGURES

Figure Page

1 Emblem of Ankara Fen Lycée 6

2 Workflow of Meta-analysis 43

3 Flowchart of the Meta-analysis Conducted 48

4 Steps Followed in Statistical Analysis 56

5 The Forest Plot for Included Studies 60

6 Baujat Plot 66

7 GOSH Plot 68

8 Summary of the Investigation of Possible Outliers and Influential Cases .. 69

9 Funnel Plot of all Included Studies 77

10 Trim and Fill Funnel Plot for all Studies 80

11 Trim and Fill Funnel Plot for Studies except the Possible Outliers 81

CHAPTER 1: INTRODUCTION

Introduction

This research is about a meta-analysis on enrichment programs conducted on gifted students with its primary concern on their academic achievement. In the background section, I present a general discussion on modern schooling and give information regarding the main stakeholders of mass education. Then, I progress to the problem section, where I presented the rationale for the need of enrichment programs as they are related to gifted education justified through two main approaches. The first was the human assets theory, and the latter was based on the constructivist paradigm. After that, I have presented the purpose and significance of this study. Finally, I mentioned the study's limitations and then concluded by defining key terms.

Background

General Discussion on Modern Schooling

Both public and private schools accomplish several important missions. While a few millennia ago, schools were private and selective establishments that only accepted a minority of people, they have now become an integral part of what can be defined as contemporary mass education serving the general public (Boli et al., 1985, p. 145). Modern life has imposed several important missions on schools. These missions can be understood or rationalized by defining two main categories: the broader goals of mass education and students' personal interests.

From the broader perspective, we recognize the main stakeholders interested in schooling to be establishments such as the society, the government, businesses,

and so on. To some degree, abstract phenomena such as culture should also be realized as the main factor in this complex equation because the continuity of a nation's culture depends on mass education through schooling. No doubt, schools have been given the important task of transmitting the culture and values of the current and past generations (Battiste, 2011, p.193). Only through mass education and schooling, it is now possible, or at least it seems as such, to transcend these values and practices accumulated over centuries by the older generations. Oddly enough, societal transformation is also a major responsibility of schooling (Desjardins, 2015, p.239). Cultural continuity or preservation and societal transformation might seem to be two contradictory concepts, but both are inherited in the culture and nature of progress itself. Nations, through schooling, pride themselves on progress and general advancement. After all, it is being theorized by many politicians that schooling leads to "social transformation," which is closely related to economic growth, technological progress, and bettering humankind (Desjardins, 2015, p.241; Stromquist, 2002, p.61). Therefore, schooling through mass education is constantly under scrutiny by various stakeholders, obviously with different concerns. The major players in this endeavor are the government, military, business, and academia, all with different agendas. For example, the government's role in education is primary concerned around citizenship issues. This concern is reflected in the must requirement; the law states that each child receives compulsory education of a minimum amount of 12 years in Turkey. Besides the government, the military and businesses have major concerns and expectations from schooling. From their perspective, a nation's livelihood and future depend on schooling. We see this concern and can provide a platitude of evidence from the past.

Gifted Education in the Past

Davis et al. (2010) cites Meyer (1965) to describe how researchers considered education in ancient Sparta as one of the early examples of exclusive education for gifted students. However, giftedness was not defined as it is defined these days. The critical selection criterion was the military skills of the individuals. Also, in ancient Greek, upper-class, free individuals continue their education in private schools. Other than Plato's Academy, these are paid institutions that accept only boys as students (Davis et al., 2010, p.4). In early China, giftedness and talent were recognized and cultivated more similarly to today's practices. Other than the military skills or social class individuals belong to, people's several abilities were considered giftedness, such as literary ability, leadership, imagination, and reasoning. Confucius, who lived around 500 B.C., is credited the Chinese at that time recognized that while all children should have access to education, each child should have an education tailored to their abilities (Davis et al., 2010, p.4). This idea is closely related to today's constructivist approach. There were other implications from all around the world. Some defined the students as gifted, while others specify the talents of the students that take exclusive education but do not label the students as gifted.

Major Contributors to the Gifted Education Field

Along with the gifted education in ancient civilizations, it is worth mentioning important researchers. Some of the first major contributors to the field were Francis Galton, Alfred Binet, Lewis Terman, and Leta Hollingworth (Davis et al., 2010, p. 5).

Galton was considered one of the first researchers who worked on intelligence. He tried measuring intelligence in his laboratory in the South

Kensington Museum of London. He focused on the individuals' reaction times to stimuli and their auditory and visual sensory discrimination abilities. He concluded that intelligence is positively correlated with the discriminative capacities of the people (Brody, 2000, p.16; Davis, et al., 2010, p.5). After Galton, Binet tried several tests to measure intelligence that failed, some of which focused on the elementary cognitive processes such as reaction time to sound, hand squeezing strength, and the amount of pressure on the forehead that causes pain. After those unsuccessful attempts, he found some relevance when he tried to link intelligence with more complex mental processes like; attention, memory, judgment, reasoning, and comprehension (Brody, 2000, p.17; Davis et al., 2010, p.5). Another important contributor Lewis Madison Terman is known as the "father of the gifted education movement." The first major contribution that helped him earn this title is adapting the Binet-Simon intelligence tests to American concepts and the broad research on the identification and longitudinal study of 1528 gifted children. He had many inferences about those students. Some of these are that parents and family values impact the success of gifted students and the positive effect of acceleration. However, identifying the students considering only the IQ values is criticized as limiting since Terman may not detect artistic or creative genius students. Although not as famous as Terman in her lifetime, another researcher who contributed to the field was Hollingworth. After her contributions were realized, she started to be known as the "founder of the gifted education field". She worked on the topics that are still a debate in today's gifted and talented education. She focused on how gifted and talented students are identified, whether acceleration or enrichment is more effective, and whether educating those students separately or in a regular classroom is better. In their experiment, they put effort into finding out gifted students' needs.

They started a school called Speyer, where gifted students were taught for three years. She made special efforts to identify students who were culturally and economically diverse. She observed those children during their high-school, college, and after-college years. The students learned everything at their own pace in accelerated and enriched content. The Speyer school caught the attention of not only the public but also educators from different countries. Her findings showed that both acceleration and enrichment are vital in gifted education as gifted students both need additional work and rapid progress (Davis et al., 2010, pp.7-8; Klein, 2000).

Effects of Sputnik's Launch on Gifted Education

It is equally important to mention the historical events that have hugely impact on gifted and talented education. In more modern history, a well-known example that affected education in America and Western schooling in general is, "Sputnik," (Davis et al., 2010, p.8). The Sputnik was the first artificial satellite launched and placed in orbit on 4 October 1957 by the Soviet Union. The initial reflex was immediately felt in education and schooling, not only within the USA but, in fact, beyond the borders of the USA (Sak et al., 2015, p.112; Wissehr et al., 2011, p.368). This crisis was a major phenomenon in the "Cold War" that initiated the creation of the National Aeronautics and Space Administration (NASA) and the space race between the USA and USSR.

The North Atlantic Treaty Organization (NATO) members, including Turkey, felt an immediate impact. The military and businesses were the stakeholders that advocated for schools to train primarily engineers with the promise to combat the "Cold War." Science Lycée (Fen Lisesi) is an example from Turkey (Sak et al., 2015, p.111). One should point out that the Turkish schooling system lies on an equality premise. Pursuant to section two of the law numbered 1739, all students,

despite their backgrounds, are promised equal public schooling (Milli Eğitim Temel Kanunu, 1973). Although a public school, despite the equality premise, the primary goal of this now selective “Science Lycée” was to choose the brightest and most promising Turkish students to be educated in science majors that could lead to well-educated engineers. The purpose was to develop an arsenal of well-educated citizens that could fulfill the societal needs to combat the “Cold War” and thus lead the nation to prosperity (Sak et al., 2015, p.111). One must understand the actual thread the “Sputnik” event possessed. The Soviet Union proved they could launch rockets armed with nuclear weapons in Western Europe and even North America. At that time nuclear threat was real.

Figure 1

Emblem of Ankara Fen Lycée

Note. The emblem of Turkey's first established science high school. Reprinted from Ankara Fen Lisesi, In *Wikipedia*, https://tr.wikipedia.org/wiki/Ankara_Fen_Lisesi

Figure 1 shows the emblem of the first established Science High school. All other high schools based after that in Turkey have a similar design. It resembles the Bohr Atom model in which the electrons orbit around the nucleus of an atom. This can be considered as another evidence manifested in the “real” perceived threat of nuclear war and its consequences on schooling. This is only one example. While the context might differ, there are many more examples, like the establishment of

Türkiye Bilim ve Teknolojik Araştırma Kurumu (TÜBİTAK) (Sak et al., 2015, p.112). It is beyond the scope and my intention for this study to account for all the historical events associated with the effect of the demands presented by the three stakeholders. My examples above should suffice to make my case for these stakeholders' possible effects and outcomes in schooling.

Problem

Bright and relentlessly hard-working students are seen as another precious "national" resource (Wallace, 1989). Terms such as human capital, human assets, stock of knowledge, and 19 more have been used in social sciences and the field of economics to describe this phenomenon. This concept has even been a part of the recent Nobel Prize for Economics award in 2018. Paul Romer, who was jointly awarded, is the founder of this concept of understanding and explaining economic growth. This concept (i.e., human asset) can be applied to contemporary growth theories as an important economic growth factor. Furthermore, the research conducted by the scientific community continuously shows the relevance of education for the economic welfare of people (Hanushek et al., 2008). Not surprisingly then, human asset or capital management becomes a key concept, especially the capacity to develop talent, the capacity to deploy talent, and the capacity to draw talent from elsewhere. Collectively, these three capacities form the backbone of any country's human capital competitiveness (Crook et al., 2011). Needless to state that solid evidence exists that nations that possess and cultivate their human assets outperform other countries deprived of human assets. Lately, in Turkey, we see that this very same concept is in place. The governments recognize that human assets are now one of the essential foundations and thus, serve as a backbone of human/social and economic development. For example, conscious

efforts are made through official calls to “recall” or attract Turkish scientists that work abroad (Sönmez, 2019). Appealing financial promises are made for those who would commit to return and work in Turkey. When analyzing the educational system through the lens of the “human assets” theory, we see two distinct approaches.

The first one is a structure around what can be defined as the liberal arts approach. This type of system draws its promise from the fact that talent can only reach its fullest potential through the provision of experiences provided by a multitude of interdisciplinary subjects. The students are allowed to pursue their education through numerous courses in the hopes of finding their talent. Of course, such an educational system needs many resources (Godwin & Altbach, 2016). Thus, it is not very feasible in developing countries. So, it is not a common practice used.

The second approach could be described as a filtering or screening approach. As the name implies, the students are elected while pursuing their education. This selective process seems to be more prevalent among the two approaches. It is less resource-dependent and tires to select the “correct” students. The selection and filtering process can start as early as the end of elementary school and can be multi-staged. Many smaller European countries do employ such a selective schooling system. For example, Switzerland provides two different schooling routes for students after finishing and graduating from primary school (6th grade). Based on their graduation grades, teachers’ evaluations, and achievement tests the students are placed into two different leveled schools that apply different curricula. And then, after finishing secondary school, the students are again placed in other high schools that follow respectively different curricula (Neuenschwander & Garrett, 2008). Although the Turkish educational system has a similar filtering process that provides

alternative schooling routes for students with different abilities, it is not working with the same efficiency as its European counterparts. Furthermore, Turkey employs the filtering process first at the end of middle school through national high-stake exam and then when transitioning to higher education (Atılgan, 2017). Sadly, the system allows for bypass during the filtering. Students who are not doing particularly well in these selective high-stake exams are bypassing the purported route by signing up for private schools if their families can afford it. So, in essence, the selective system provides loop wholes. Another problem with the selection process in Turkey is that it is based on ranking (İş Güzel, 2013, p.12). Since ranking is not a score, it cannot be a standardized measure. This leads to a multitude of problems in the selection process. For example, a highly ranked student in a densely populated district could end up at a lower-level school than a student in a less dense region of Turkey. Hence, it is not uncommon for some families to relocate to other cities to increase their children's chances of attending a higher-grade school. So, placement or selection is arguably not based on an absolute normative scale. The result of this selection process is that students with an overall lower ranking can end up at schools that apply high-level academic curricula, such as the science lycée. Another negative outcome of these exams is that the community and sometimes the policymakers draw conclusions about the effectiveness of the education by looking at the results of these exams with previous years. However, it is not possible to compare and contrast the results as one cannot know whether the exams are parallel in terms of difficulty, which in fact is a problem of content validity (İş Güzel, 2013, p.12).

So far, I have provided two concrete examples of schooling needs beyond the basic approach of citizenship education and the need to transcend the culture to the future generation. As seen from the examples above, stakeholders are mainly

concerned with the efficiency of student outcomes. Through this process, the human asset quality concern for the general public welfare is rationalized through selective schooling. These concerns have sparked many highly political (and, to an extent, radical) discussions in many nations. For example, *"The Bell Curve,"* a book published by psychologist Richard J. Herrnstein and political scientist Charles Murray (1996) makes an argument on these very same issues and advocates for the "cognitive elite." They provide an "intellectual" argument on who should receive an education and at what academic level. They question the resource allocations and make the case that some people inherently are limited; thus, resources should be spent based on a rationale such as race and intelligence. A plethora of similar arguments exists in many other nations. The human assets point of view is understandable, highly controversial, and politically charged in nature. Despite these controversial debates, we still see nations apply different curricula to the selected population of students.

The second major domain of the problem lies within the learning theory that constructivism promotes. The constructivist theory has shown that students have different learning needs and ways. School administrators and teachers have embraced this reality. Those days are long gone when the traditional teaching paradigm treated all students as unitary or a tabula rasa (blank slate) as advocated by John Locke (Bada & Olusegun, 2015; Pritchard & Woolard, 2010, p.48). We now know that students come with their own/unique concepts and cognitive schemata that need to be considered in the learning and teaching process. Concepts such as differentiation, multiple intelligence, resilience, student-centered learning, active learning, lived experiences, intrinsic and extrinsic motivation, and so on are all part of making meaningful sense of student needs in their learning processes. All of these

concepts, regardless of age and level, are contemporary domains of research conducted in academia – the fourth stakeholder in mass education and schooling mentioned earlier in the introduction part of this chapter. The student is now at the center, and we theorize that the learners actively construct knowledge to reach a meaningful understanding (Pritchard & Woolard, 2010, p.47). As students experience the world, our task as teachers has become to provide them with the opportunity to reflect upon their experiences (Pritchard & Woolard, 2010, pp.48-49). Students build their own representations through cognitive processes such as assimilation and accommodation by incorporating new information into their pre-existing schemas (Pritchard & Woolard, 2010, p.49). Classical teaching values the transmission of information, but constructivists claim that knowledge cannot directly be imparted to students. Only experiences can facilitate students to construct and make meaning of their knowledge. Therefore, the goal of teaching is to consider students' individuality and design these teaching activities accordingly. As a result of this new paradigm, students' needs, and requirements are now taken into serious consideration in schooling. Teachers and school administrators make every effort to fulfill students' needs. This holds for a wide ability spectrum of students, including gifted students.

A simplistic or traditional approach toward the academic needs of gifted students was in the old paradigm, as explained above, just to ignore them. After all, these students are by definition above average and should or could deal with any academic issue without the need for teacher intervention. However, recently we have seen an increase in research conducted on various issues on gifted students, including their needs partially fulfilled through enriched programs that claim to enhance or provide a more meaningful learning environment for these cohorts of students. These

publications work towards a better understanding of these many different enrichment theories. Moreover, this type of approach was reported to increase students' efforts, enjoyments, and performance with the bonus of the integration of advanced level learning experiences and thinking skills into their schooling experiences (Kuykendall, 2020; Preckel et al., 2019).

Therefore, it is worthwhile to investigate the claims through a meta-analysis. As a result of such a study, we would be able to see if the advocated use and extension of some types of enrichment programs for the talent indeed is a real claim.

Purpose

The purpose of the study was to conduct a meta-analysis on the studies that primarily dealt with research on the outcomes of enrichment programs on the academic achievement of gifted and talented students by estimating the overall effect size (OES). Utilizing a quantitative research approach, this study systematically reviewed the published research. A pooled effect size value was calculated to show the direction and the strength of the relationship between the enrichment programs and the academic achievement of gifted and talented students. The rigorous process of selecting the relevant peer-reviewed published research created the fundamental data for the meta-analysis. More specifically, the study sample was the academic articles published between 2013 and 2022. The study used the standardized mean difference (SMD) as the effect size type to examine data to identify and consolidate the outcome.

Research Question

How do enrichment programs affect the academic achievement of gifted and talented students?

To understand the findings clearer following hypotheses were tested statistically in the subgroup analyses:

Null hypothesis 1: There is no difference between the effect size of the studies that have a small sample size (n<100) and a large sample size (n>100).

Null hypothesis 2: There is no difference between the study's effect sizes with different grade levels of students as their sample.

Null hypothesis 3: There is no difference between the effect size of the studies that are written in English and Turkish.

Null hypothesis 4: There is no difference between the effect size of the studies that are designed as one group pre-post test and control treatment groups.

Null hypothesis 5: There is no difference between the effect size of the studies that are published in the journal and that are doctoral dissertations that are not published.

Null hypothesis 6: There is no difference between the effect sizes of the studies that are using the same pre-test and post-test and different pre-tests and post-tests.

Null hypothesis 7: There is no difference between the effect sizes of the studies that are investigating different subject areas.

Significance

Enrichment programs designed to focus on the gifted and talented students' interests are argued to significantly increase students' learning experiences. For example, it is found by Schäfers and Wegner (2020) that it helps these students to identify their interests in a more meaningful way and pursue them later in secondary school and college, leading to the selection of majors and careers. Their results show that students taking enrichment are three times more likely to pursue STEM-related

jobs than their national counterparts. So, from the point of the human assets theory, this would have resulted in the desired outcome—efficient use of the resources. Baum et al. (1995) showed the same positive outcome in academic achievements with the added benefit that the students were better able to identify their strengths. Longitudinal research on the use of this approach confirms the outcomes. For example, a very recent publication by Brandon et al. (2021) showed that autonomy for the creativity of talented university students was directly related to the enrichment opportunities the programs provided.

A second important outcome as far as the human assets theory is concerned is the strong positive relationship between the early interest and its stability. What that means is that students' interest shows a significant correlation with their subsequent upcoming school plan decisions, future profession and career choices, long-term goal persuasions, and overall self-regulation. These outcomes are auspicious when taking the Turkish context into account. As earlier described in this chapter, the filtering process in place in Turkey is not an efficient selective school system because it allows for students to be misplaced. As a result, students that could be considered gifted or talented may well attend lower-level schools or vice versa. Enrichment programs, if proven to be significant, might be an alternative to implementation to increase the efficiency of student assets.

A recent meta-analysis using 13 articles published between 1985 and 2014 by Kim (2016) showed that enrichment programs positively affect gifted and talented students academically (g=0.96, 95% CI [0.64, 1.30]). However, the literature has grown so much that we have the chance to conduct a meta-analysis using the data from studies published between 2013 and 2021.

Hence, a meta-analysis will further our understanding and be a significant add-on to the scientific literature for future policymakers on this efficiency issue.

The notion of curricular diversity in education is a recent one. In Turkey, we have several different curricula, the former minister of education announced that 51 different curricula were revised in 2018 (MEB, 2017). Although they are different from each other, they are centralized which, does not provide much room for change. We have recently seen new attempts by the Turkish Ministry of Education (MoNE) under the name of "Science and Art Centers (Bilim ve Sanat Merkezleri)." The target is gifted and talented students. As of this writing, 182 centers were fully operational, 63 thousand students were enrolled, and 2223 teachers are working in these centers. The efforts over the past ten years of MoNE showed a steady increase in enrolled student numbers. MoNE reported within these ten years, the number of centers increased from about 60 to 182 (MEB, 2020). Therefore, this study is a timely scholarly work that will contribute to the much-needed body of research.

Limitations

This study has some limitations that the readers should be cautious about when interpreting the results. Firstly, there are several problems that one can encounter in any meta-analysis. These are: apples and oranges problem, garbage in garbage out problem, file drawer problem, and researcher agenda problem (Harrer et al., 2022, pp. 9-10).

The apples and oranges problem suggests that even though the researcher used very strict inclusion criteria, there will always be minor or significant variations in the included sample. Estimating cumulated studies is always possible from a statistical perspective, but it loses significance when studies do not share the characteristics necessary to respond to a particular research question.

The garbage in garbage out problem refers to the impact of the studies' quality included in a meta-analysis. If the studies included are biased, so are the meta-analysis' results.

The file-drawer problem arises from the missing relevant studies in the meta-analysis. The negative or disappointing results are not published like the innovative findings, which creates publication bias. In this research, the Google Scholar database was screened in detail to find relevant grey literature.

Last but not least, while conducting meta-analyses, the researchers had to decide many things, such as which keywords to search and which articles to include or exclude. So, meta-analysts' decisions may influence the findings. This problem is called the researcher agenda problem. In this thesis, the researcher consulted the advisor in the critical parts to minimize this problem. For instance, both the researcher and the advisor read full-text articles separately to decide whether it is appropriate to include or exclude them, then they discussed them together. However, at the beginning of the screening, the researcher decided which articles to include or exclude that can be considered a part of the researchers' agenda problem.

Also, all the studies included in this study are written either in English or Turkish. The reason is that the researcher is not proficient enough to understand any other language. As a result, this study may have a language bias (Harrer et al., 2022, p. 229).

In addition to these, the definition of giftedness varies from study to study. The definition of "gifted and talented students" used in this study is an inherent limitation.

Definition of Key Terms

Academic Achievement: Although academic achievement can be defined broader, in this study, it is restricted to the content-area achievement as measured in language, reading, mathematics, science, or social science.

Acceleration: "Any strategy that results in advanced placement or potential credit" (Davis et al., 2010, p.128).

Between Study Heterogeneity: "The extent to which true effect sizes varies within a meta-analysis" (Harrer et al., 2022, p. 140).

Enrichment: "Strategies that supplement or go beyond standard grade-level work, but do not result in advanced placement or potential credit" (Davis et al., 2010, p.128).

Fixed-Effect Model: A statistical model that is used to explain the reasons why and how observed effect sizes differ. It assumes that all the studies have the same true effect size. The variation between the observed effect sizes results only from the sampling error (Harrer et al., 2022, pp. 94-95).

Gifted and Talented Students: There are various definitions of gifted students. In this study its definition is very broad. This study involves all the studies that are identified gifted students by using standardized tests, IQ tests, students' GPA, considering the opinions of the teachers or families or a combination of these.

Effect Size: "Effect size is a metric quantifying the relationship between two entities. It captures the direction and magnitude of this relationship" (Harrer et al., 2022, p. 54).

Meta-Analysis: "The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding" (Glass, 1976, p.3).

Overall (Pooled) Effect Size: Effect size of the studies included in the meta-analysis.

Pull-out Program: "A pull-out program is one in which the students leave their regular classroom at regularly scheduled times for special programming" (Wallace, 1989, p.81).

Random-Effects Model: A statistical model that is used to explain the reasons why and how observed effect sizes differ. It assumes that not all the studies have the same true effect size. There are other sources of variance other than sampling error. So, the overall effect size is the mean distribution of the true effect sizes (Harrer et al., 2022, pp. 99-100).

Robustness: The findings of the study do not heavily depend on one study (Harrer et al., 2022, p. 153).

Standard Error: "The standard deviation of the sampling distribution" (Harrer et al., 2022, p. 56).

Standardized Mean Difference (SMD): Type of effect size measured by: "difference in means between two independent groups, standardized by the pooled deviation (Harrer et al., 2022, p. 65)."

Statistical Heterogeneity: "A quantifiable property, influenced by the spread and precision of the effect size estimates included in a meta-analysis" (Harrer et al., 2022, p. 140).

True Effect Size: The theoretical effect size of the intervention.

CHAPTER 2: LITERATURE REVIEW

Introduction

This chapter aimed to review the research aligned with this study's purpose. The aim of this study was to estimate an overall effect size of the impact of enrichment programs on the academic achievement of gifted students. So, this chapter starts with a broad perspective on the literature on gifted and talented students as it relates to formal education. Subsequently, enrichment programs will be discussed. After that, the focus of the discussion switches to the understanding of academic achievement. Lastly, meta-analysis and its implications in gifted education will be mentioned.

Gifted and Talented Education

In the twenty-first century, constructivism has become one of the leading philosophies in education. With this philosophy, the emphasis of the education becomes students (Krahenbuhl, 2016). It should not come as a surprise then to see teachers, administrators, and researchers trying to understand what and how to teach these students that stand out from the crowd. Many schools have already tailored their education towards students, and we see this for gifted and talented students as well (Reis & Renzulli, 2010). The aim to tailor and serve the needs of gifted students brings unique challenges to the field of education and related entities. The following paragraphs will point out some of the problems associated with gifted and talented students as it relates to education.

One of the problems perceived in this area is the operational definition of gifted and talented students (Kuykendall, 2020, p.7). As with any scientific term or

terms used in a scientific context, giftedness has various definitions. Generally, this term has been associated with IQ and academic achievement test results. A natural tendency by the readers is the presumption of high cognitive function executions by the person labeled as gifted or talented. This type of identification is called the "Gifted Child Paradigm." This approach is the earliest and most traditional one used in the identification process. In this approach, being gifted is a fixed, inherited trait (Dai & Coleman, 2005, pp.377-378). However, some problems arise when defining giftedness only as an IQ concept (Getzels & Jackson, 1958). For example, one of the problems is that it limits the identification of gifted students. Many questions like what the threshold IQ score should be or how we should assess IQ arise with this problem. Since the mid-80s, this definition has been challenged by a new approach called the "Talent Development Paradigm." The advocators of this new approach believe that giftedness is not fixed but rather a trait that can be developed (Dai & Coleman, 2005, p.378). Another approach, the "Differentiation Paradigm," also considers giftedness dynamic. Its focus is more on the needs of gifted and talented students. Rather than the long-time effects of gifted education, the supporters of this idea focus on what students need now (Dai & Chen, 2021, pp.180-181).

One can also observe that different countries, states, or even schools define giftedness differently. For now, there is still no consensus on the definition. A lack of a common definition poses a legislative problem for school districts. For example, states have different definitions of giftedness when it comes to identify students. In Alaska, "Gifted means exhibiting outstanding intellect, ability, or creative talent." In Louisiana, "Gifted children and youth are students who demonstrate abilities that give evidence of high performance in academic and intellectual aptitude." These types of definitions then create operational problems. Nevertheless, gifted and

talented students, no matter how defined are present. Furthermore, schools (even governments) create programs for the potential benefit of the nation and students.

There is a secondary problem as well. What to do with those students identified or selected as gifted and talented? Like its' definitions, there is no consensus on a clear purpose of gifted education. Researchers to point out whatever the purpose, at the end society as well as the students will benefit. One of the supporters of this view states that many gifted students and their needs are not identified. Gifted students may find school intolerable and boring as they lack the challenge they need. Others tolerate school and fulfill their needs outside the school time. This may result in harm to both the students' academic life and the improvement of society (Davis et al., 2010; Plucker & Callahan, 2020). In line with this view, Renzulli emphasizes that gifted and talented students need to be taken into account by the educational systems so that the nations can gain - "the major purpose of gifted education is to increase the world's reservoir of creative and productive people" (2021). On the other hand, some educators emphasize the need of the student. Gifted and talented students need to have tailored education considering their needs grounded in equity and focus on individual self-actualization (Roeper, 1990, as cited in Grant et al., 1999).

Besides the different ideas on its definition and purpose, there are criticisms against gifted education. Plucker and Callahan (2020) listed three of them. Firstly, these programs are criticized for being elitist. The advocators of this criticism argue that these students already have the chance to cover everything themselves, and hence, the special services and efforts offered to gifted students are elitist in nature (Plucker & Callahan, 2020; Sapon-Shevin, 1996 as cited in Borland, 2003; Wells & Serna, 1996). Another opposing argument is an overrepresentation of white, high-

income students in the identification process. This particular argument comes back to the definition and hence selection process of gifted and talented students (Sapon-Shevin, 1996 as cited in Borland, 2003; Kitano, 1998 as cited in Borland, 2003 pp. 160-165). They believe this process is biased and it favors white, privileged students. However, gifted education is widely accepted, and different implications are applied worldwide.

In addition to the disagreement in the definitions, purposes, and gifted education, there is also a struggle with the programs or strategies that are purported to satisfy these students' needs. The two common strategies employed are called *acceleration* and *enrichment*. Acceleration is defined as "any strategy that results in advanced placement or potential credit" (Davis et al., 2010, p.128). Some examples of acceleration are early admissions, grade skipping, and curriculum compacting. Enrichment refers to "strategies that supplement or go beyond standard grade-level work, but do not result in advanced placement or potential credit" (Davis et al., 2010). In a well-rounded program, both of these strategies can take part. However, there are other programs focusing only the enrichment or acceleration. This study focuses on the programs that contain enrichment strategies. So, enrichment programs and examples of possible strategies that are applied in the included articles will be discussed in detail in the following parts.

Enrichment Programs

Enrichment programs increase the depth and breadth of traditional subject matter (Roberts, Inman, & Robins, 2022, p.203). They can be considered as a form of talent development. Much of the groundbreaking research on talent development models come from cumulative works of Joseph Renzulli, John Feldhusen, and Donald Treffinger (Kuykendall, 2020, p.27).

Enrichment Models

Although there are different models, such as The Purdue Three-Stage Enrichment Model and Levels of Service (LoS), the majority of the current enrichment approaches depend heavily on Renzulli's enrichment triad model (Worrell et al., 2019). The scope of this research is not to investigate these models in-depth, so the Purdue Three Stage Enrichment and Level of Service models will be discussed briefly. However, Enrichment Triad Model will be mentioned more detailed.

Enrichment Triad Model

The enrichment triad model developed by Renzulli contains three types of activities. These are General Exploratory Activities, Group Training Activities, and Individual or Small Group Investigations, also called Type 1, Type 2, and Type 3 activities sequentially. In the updated version of the Enrichment Triad Model, Renzulli (2021) mentions the importance of the arrows as they make the model dynamic. Implementing these three types independently is not as effective as implementing them altogether. Type 1 and Type 2 activities also contribute to the identification process as it enables the teachers to observe that enthusiastic students have high interest and motivation for tasks. These students may not be identified through standardized tests.

Type 1 activities, which include all students, focus on students' personal interests. The activities in this part aim to motivate students to act on their interests. Some example activities used for this purpose are inviting a guest speaker, organizing a debate, and showing videos. The successful Type 1 activities are dynamic and invite the students to follow up activities.

Type 2 activities aim to foster students' skills such as critical thinking, problem-solving, etc. Activities in Type 2 are not decided a priori, but they are planned after students attend Type 1 activities and find that they have an interest in a specific topic. During these activities, students develop the skills referred to in the taxonomy of cognitive and affective processes.

Type 3 activities allow investigating a real-world problem. There are four characteristics of real problems. These are personalization of interest, use of authentic methodology, no existing solution or right answer, designed to have an impact on an audience other than or in addition to the teacher. The aim of these activities is for students to provide opportunities for students to think, feel, and do like adult professionals at a junior level. As seen in the first two types, they are designed for all the students, whereas the third type is suitable for highly motivated students with the appropriate skills to proceed.

The Schoolwide Enrichment Model (SEM)

Reis and Renzulli (2010) denotes the rationale to develop the school wide enrichment model is to "encourage and develop creative productivity in young people". They argue that a shift is needed in the concept of gifted and rather making it static as being gifted people should focus on the development of the gifted and creative behaviors. They stated that the enrichment triad model lies in the core of this model. Using the Revolving Door Identification Model, they increase the chance of the students to continue in type three activities. In this model students have the chance to receive enrichment in different levels rather than excluding from the program. SEM not only involves enrichment activities, but it also includes acceleration, differentiation and curriculum compacting. It provides opportunity to develop creative-productive giftedness along with academic giftedness.

Three Stage Enrichment Model

Influenced by both Renzulli's Enrichment Triad Model and Stanley's research on the need and effectiveness of acceleration, Feldhusen, Linden, and Ames developed a new model for gifted and talented students (Feldhusen & Kolloff, 1988). Originally, the model was developed for the use of college-level students, and then it was adapted for elementary levels. This curriculum framework aims is to master the basic concepts and develop higher-level thinking skills (Moon et al., 1993). In his book, where he explains the course design of the model, Feldhusen (1980) advised dividing the instruction into six to eight units with a common theme or unity. For each of these units, three steps that give the name to the model are developed. The steps are summarized in the Table 1 below.

Table 1

Three Stage Model

Stage 1 Self Instructional Guide (SIG)	Stage 2 Group Instructional Guide (GIG)	Stage 3 Procedures for Individual Projects (PIP)
The aim is to learn the basic concepts through individualized instruction.	Small-group problem-solving activities are planned to foster higher cognitive skills.	Students individually work and transfer their knowledge to real-world problems.

Levels of Service Model

Like the Schoolwide Enrichment Model, Level of Service Model is relatively a new approach that focuses on providing a greater number of students with enrichment opportunities (Kuykendall, 2020, p. 28). There are four levels which are interrelated to design programs for talent development (Treffinger & Selby, 2009). The levels are similar to enrichment triad models in some ways. For instance, both models start including broader number of students and as the level of complexity increases, the number of students' decrease. Treffinger et al. (2008) mentioned the

first level includes more basic approaches to provide foundational skills. In this step all the students benefit from enrichment opportunities. Then in the next level, many students continue exploring their interests. After that, some students go further to fulfill their needs. Lastly, few students proceed to blossom their expertise with a highly individualized tasks.

The common strategies used in enrichment generally involve combination of applications such as independent studies, learning centers, field trips, Saturday and after-school programs, academic competitions, pull-out services and summer programs.

Strategies Used in Enrichment Activities

Independent Studies

Independent studies' topics can be selected considering the students' interests. This topic may change throughout the research process. Some research topics may include field or museum trips as well. The critical point is to teach students to triangulate the data from the sources they collect and evaluate the validity of the sources. Along with internet search, students can conduct scientific research to explore a problem that does not have a predetermined result. In this type of project, students need to collect and analyze their findings. Other than these, students can conduct projects related to art, drama, or creative writing. In every project, the students must create a product or a performance and present it to the audience (Davis et al., 2010, p.16).

Learning Centers

Learning centers involve materials, games, and tools for a specific subject area. For instance, a science center may contain science activities, recordings, and videos. An alternative way to learning centers is outdoor learning centers. Unlike traditional

classrooms, where they read and discuss the concepts that lack change, outdoor learning centers provide students with complex real-world settings (Pfouts & Schultz, 2003).

Saturday and After School Programs

Saturday programs do not have a strict curricular agenda like schools. They can have special classes about different topics such as astronomy, botany, or robotics. The instructor does not need to be a teacher, but they can be an expert, a graduate student related to the field (Feldhusen & Ruckman 1988).

Academic Competitions

As discussed earlier, there were many changes in gifted education after Sputnik's launch. Academic competitions are one of them. After the Sputnik’s launch, Edwin Teller, one of the science advisors of the president promotes implementing academic competitions to Eisenhower. By doing so, students will engage in science and technical areas in earlier ages which is desirable in human assets theory. The assumptions of the benefits of these competitions are early identification of gifted students, encouraging schools to differentiate their curriculum, attracting, and motivating participants, and developing society as talented individuals are developed on the way (Campbell et al., 2000).

Academic Achievement

There are many variables affecting students’ academic achievement. That’s why there exists more than one definition for academic achievement. In general, it is closely related to skills and competencies that allow students to be successful in school and society. A recent systematic review of meta-analyses investigated the factors affecting academic achievement. From 169 meta-analyses, they explored 254 variables in nine categories. These categories were psychological, learning theories

and teaching strategies, family, teacher, school, educational technology, special education, and violence (Koçak et al., 2021, p.455). Even these numbers can show the readers how complex the academic achievement concept is. Concepts such as success, achievement, and particularly academic achievement are difficult to define and measure as they are dependent on many variables. As a result of this, researchers sometimes narrow the definition to investigate achievement (Lindholm & Borsato, 2006, p.176). This study focuses specifically on the school subjects such as mathematics, science, language, social studies, and GPA.

Meta-analysis

The idea of meta-analysis goes back as far as 1930s. Rosenthal (1991) provides an account of three types of research what could be classified as meta-analytic approaches. These types of research were not named explicitly meta-analysis. Similar to Rosenthal, Glass (1976) gave many examples of studies in which the researchers tried to organize and generalize the findings on different topics. Another early example comes from Astin and Ross (1960). Their research was based on a quantitative approach on studies exploring the effects of glutamic acid on the intelligence of retarded children. As research grows, there is an obvious need for summarizing and providing some sort of a generalization based on the quantitative research paradigm (Glass, 1976).

Although, meta-analysis is closely related to systematic literature review, both have a systematic selection process in which the researcher spends a lot of time finding the relevant literature, it is different. While systematic literature reviews are based on qualitative paradigm, meta-analysis is very different in nature. It is purely quantitative and follows statistical theories (Harrer et al., 2022, pp. 4-5).

Advantages of Meta-analysis

The growing literature creates the demand for combining the results. One way is literature reviews which are mostly qualitative. Another one is meta-analysis. It gives the opportunity to statistically generalize the studies conducted on the same research questions before (Fagard et al., 1996).

An advantage of meta-analysis is that the effect of the small trials that may not be significant on their own becomes clear in pooled effect size. Generally, a clinical study is not sufficient to estimate the true effect size. The advantage of meta-analysis is that the sample size is bigger than clinical trials. As the sample size increases, so is the statistical power. By doing so, the estimate becomes closer to the effect size. They are especially useful when the studies are so small to give reliable estimates to the interventions' effect size (West, 1993).

Another advantage is that unlike systematic review, meta-analysis pools the results in a more scientific way. Fagard, Staessen and Thijs stated that systematic reviews have been unsystematic, and they lack scientific rigor (1996).

Also, Cooper and Rosenthal (1980) highlight the importance of accountability. A meta-analysis provides the reader with the selection criteria that are then replicable, resulting objectivity. Because of the openness of the data sets, interested parties can use the validation process and data set to confirm the analysis and statistical procedures. It is an advantage as the traditional reviews have been criticized since they do not allow researchers to control findings as they can with meta-analysis.

Furthermore, it was found by Cooper and Rosenthal (1980) that using meta-analysis has an advantage in decreasing the type two errors. They conducted a study with graduate students and faculty members. The participants were asked to perform

statistical analysis using the same data set. The aim was to detect the differences, if there exists any, from the conclusions they draw using traditional or statistical methods. The results of their study showed that among the thirty-two graduate students and nine professors, 68% of the statistical reviewers considered rejecting the null hypothesis. On the other hand, only 27% of the traditional reviewers do so. However, there was a clear relationship between the variables in the dataset given to the researchers. Most traditional reviewers failed to reject the null hypothesis, which was in fact false, leading to a type2 error. This shows a higher chance of making type2 error when applying traditional methods (Rosenthal, 1991, p.12).

Disadvantage of Meta-analysis

Besides the advantages of a meta-analysis, there are also some disadvantages and concerns. Researchers, for example, criticized meta-analyses for several reasons: the quality of the data obtained from the sample used in meta-analyses is a major concern. The concept of heterogeneity between the studies, and possible publication bias are also some major disadvantages (Esterhuizen & Thabane 2016). For example, the “Survival bias” holds true for publications as well. These limitations yield misleading results because we don’t have access to studies that were not published.

A concrete example “An Exercise in Mega-Silliness, Eysenck (1978) criticized Glass’ meta-analysis on the effects of psychotherapy. The reason that Eysenck criticizes is that some important data on the topic is neglected. In addition, he dwells on the quality of the material and the method. He said, “A mass of reports—good, bad, and indifferent—are fed into the computer in the hope that people will cease caring about the quality of the material on which the conclusions are based.” He concluded that he disagrees with Smith’s and Glass’ findings on the effectiveness of psychotherapy (Eysenck, 1978).

Later, Feinstein (1995) called meta-analysis the statistical alchemy for the 21st century. He uses the term alchemy because he believes that while using meta-analysis, people neglect the four established requirements of modern science. These are reproducibility, precise characterization, internal validity, and external validity. Among his criticism, he mentions that the term apples and oranges problem is not appropriate as two fruits are insufficient to express the variety of the samples in a meta-analysis (Feinstein, 1995).

Meanwhile, Liberati (1995) agrees with Feinstein in some aspects, especially on the sample quality, by saying, “The real issue, however, is to recognize that most of its limitations largely depend on the poor quality of Randomized Control Trials (RCT) and other raw material used to combine information.” Despite his opposing views, he points out the need for systematic reviews in his following sentence and says the only alternative to systematic reviews is unsystematic, implicit qualitative reviews (Liberati, 1995).

Examples of Meta-analyses on Gifted Education

Due to the advantages outweighing the disadvantages, meta-analysis studies have become widespread in many areas. The first meta-analysis in the gifted education field was conducted by Kulik and Kulik (1982). By looking at the statistics of 52 studies conducted before, they found out that high-ability students benefit more from the enrichment in ability grouping classes. The benefit is both on achievement and attitude.

A very recent article published by Oğurlu (2021) overviewed the meta-analyses on giftedness. Among the 168 studies he identified through databases, he included 22 meta-analyses in his study. His thematic analysis shows that most of the meta-analyses he found are about social-emotional development. The second most

popular topic is educational interventions. By educational interventions, he means acceleration, enrichment, and pull-out programs. All the meta-analyses on the interventions show a positive effect on students' achievement and socio-emotional outcomes. The studies related to the educational interventions for gifted students were conducted by Wallace (1989), Goldring (1990), Vaughn et al. (1991), Kent (1992), Steenbergen-Hu &Moon (2011), Kim (2016), and Snyder et al. (2019). From now onward, some of their key findings are summarized.

Firstly, Wallace (1989) list three rationales in favor of differentiate gifted students' program. Firstly, he states "by definition gifted students possess variability and potential unlike that of ordinary children". The second reason is that differentiated education helps equality of education. The last reason is that he believes gifted students are "precious natural resources". Wallace (1989) studied the effects of enrichment programs. He analyzed 20 articles that were published between the years 1931 and 1989. The results showed that the overall effect size is 0.57. In addition, he found a positive effect of enrichment programs on students' creativity and affective outcomes.

At the same year, Goldring (1990) studied the classroom organizational strategy as a factor. These were the special class model and mainstreaming model. He conducted a meta-analysis on 23 articles in his study. He found that, in general, gifted students perform well in special classrooms, especially in subjects of mathematics and science. The effect size for mathematics and science outcomes is 1.067, with a standard deviation of 0.228. However, in opposition to this, he found that the effect size for language achievement outcome was –0.369, and the standard deviation is 0.084. This means students perform better in regular classes for language

subjects. He also found that students take advantage of the special classrooms more when they get older.

Later, Steenbergen-Hu and Moon (2011) conducted research on the effects of acceleration. In addition to their own determined studies as a part of the literature review, they extended the data set by including studies from previous meta-analyses on the same topic. From the analyses of 38 articles, 15 of them investigated academic achievement effects. A ratio of 141:274 preliminary effect sizes was about academic achievement. The number of studies contributing to the combined effect size was 28. They found a positive overall effect size which was 0.180, and a relatively large standard error of 0.128. Obviously, the result was not significant. Similar to their findings on academic achievement, they found a positive, but small overall effect size which was not significant for social and emotional development.

Recently, Kim (2016) reported on a meta-analysis on the effects of enrichment programs on gifted students' academic and socio-emotional outcomes. In their study, the definition of academic achievement was extremely broad. In addition to the outcomes from subject areas, thinking processes, intelligence, and specific skills were all considered as academic achievement. The total sample size was 26, where 13 of them were for academic achievement. Findings show a positive effect both for academic (g=0.96) and socio-emotional achievement (g=0.55).

After Kim's findings, Warne (2016) published an article as he finds Kim's results unrealistically high. He conducted a statistical power analysis, which showed evidence of publication bias, and concluded that Kim's findings do not indicate that enrichment programs are effective.

Conclusion

This chapter provided information on gifted and talented students and as it related to education. More specifically, enrichment programs and their possible effects were reported on academic achievement for this type of learned. Brief discussion on giftedness was also included. The chapter also provided information regarding meta-analysis. This type of research has been recently gained popularity in social science. The advantages versus disadvantages were accounted for. The results of these meta-analyses were mostly in favor of the gifted and talented group that took intervention. However, one study's findings were negative. Albeit, there is a small chance that the findings will be in favor of the control group.

The following chapter reports on the research design, a discussion on the key concepts involved in a meta-analysis and concludes with an account on how this particular research was carried out.

CHAPTER 3: METHOD

Introduction

The purpose of the study was to conduct a meta-analysis on the studies that primarily dealt with research on the outcomes of enrichment programs on the academic achievement of gifted and talented students. In technical terms then, the primary concern of this study was to estimate the overall effect size of the enrichment programs on the academic achievement of gifted and talented students. In order to calculate an overall effect size, the meta-analysis method was employed. This chapter covered the research method and design used followed by its practical application. Table 2 below shows the timetable of this research. It involves the summary dates of major steps we took throughout the research process.

Table 2

The Timetable of Thesis

Steps	Years					
	2020-2021			2021-2022		
	Semester			Semester		
	Fall	Spring	Summer	Fall	Spring	Summer
Meetings with supervisor	X	X	X	X	X	X
Literature review		X	X	X	X	X
Getting to know R		X	X	X	X	X
Reading the book on how to conduct meta-analysis using R				X	X	X
Mockup meta-analyses					X	
Online database search				X	X	
Discussion with the advisor about including articles					X	
Data extraction					X	
Data analysis					X	X
Interpreting the results					X	X
Organizing the thesis						X
Submission of thesis						X

Research Method

Qualitative and quantitative research methodologies are the two main research paradigms. Research that employs the quantitative methodology emphasizes statistical inquiry into the observed problems. Qualitative methods are preferred when the concern is to reach out for complex information on the phenomenon through a better appreciation to the participant's background (Taylor, 2005). The qualitative paradigm describes the observed phenomenon with all its possible details through a subjective lens, as its goal is to define and understand social relations and practices in-depth (Watzlawik & Born, 2007).

In this study, the goal was to reach an overall understanding of the outcomes of enrichment programs on the academic achievement of gifted and talented students by collecting quantitative data from primary studies and then analyzing those using statistical procedures. Therefore, this study was based on a quantitative research paradigm. The research method that fits the goal of this academic endeavor was meta-analysis. The following section provides information on the meta-analysis research method and its important concepts such as effect size and its computational procedures. Then the second part of the chapter covered detailed information regarding how the meta-analysis was carried out in this research.

Meta-analysis

Glass (1976)–one of the leading authors on meta-analysis- rationalizes meta-analysis as "the analysis of analyses." What's meant with this phrase is that meta-analysis refers to the statistical procedures conducted on a large and sometimes not so large collection of individual studies that eventually form the bases to achieving an overall all understanding. Here the term *via statistical procedures* is the key and differentiates meta-analysis from a narrative review (a qualitative research

paradigm). One can conceptualize meta-analysis as any statistical procedure that collects quantitative data from secondary sources. The data sources are typically from already done research. The researcher then uses those data as an input to compute the relationship between two target variables or the effectiveness of an intervention. Hence the prefix “meta” means "more comprehensive" or "transcending”. Meta-analysis connotes a rigorous alternative to the causal, narrative discussions of literature review studies (i.e. narrative review) which typically attempt to make sense of the published literature (Gurevitch et al., 2018).

The subject of the meta-analyses is then, the statistical outcomes of scientific studies. Since meta-analysis is a statistical procedure, all statistical assumptions hold true for meta-analysis as well. The operational definition of population, the sampling method, and its limitations are important concepts. These concepts confront meta-analysis with its own challenges. For example, the concept of **the effect size** is very important in any meta-analysis. The next section provides a detailed discussion on the concept of *effect size.*

Effect Size Measures and Types

The noun word “measure” refers to “a standard unit used to express the size, amount, or degree of something”. In statistics, an “effect size” is a generic term that refers to a quantitative value representing either the strength of the relationship between two variables in a population, or a sample-based estimate of that quantity. The correlation coefficient “r” is a good example of the former, whereas the “mean” is a well-known example of the latter.

When evaluating the strength of a statistical claim such as those presented in a meta-analysis the *computation* of the “effect size” measure becomes an essential step. Lately, the practice of reporting of the effect sizes has become particularly

prominent in social science and in medical research. At this point, one might wonder if there is an expected effect size value that researchers strive for. The short answer is that there is no particular value or magic number. The interpretation of the value is contextual and mostly done in conjunction with one other important measure -the standard error.

Determination of Effect Size

The concept of effect size is a statistical measure that provides us with numerical information about the strength of quantitative research after a statistical method has been carried out on the dataset. Hence, as alluded to the previous section there are a number of different types of "effect size." By definition, meta-analysis summarizes quantitative outcomes from several studies. To perform a meta-analysis, the researchers' primary task is to determine an effect size that can be summarized across all studies. Luckily the literature on meta-analysis provides us with well established guidelines to accomplish this task; the selected effect size measure for a meta-analysis should be:

"• Comparable: It is important that the effect size measure has the same meaning across all studies. Let us take math skills as an example again. It makes no sense to pool differences between experimental and control groups in the number of points achieved on a math test when studies used different tests. Tests may, for example, vary in their level of difficulty, or in the maximum number of points that can be achieved.

• Computable: We can only use an effect size metric for our meta-analysis if it is possible to derive its numerical value from the primary study. It must be possible to calculate the effect size for all of the included studies based on their data

• Reliable: Even if it is possible to calculate an effect size for all included studies, we must also be able to pool them statistically. To use some metric in meta-analyses, it must be at least possible to calculate the standard error (see next chapter). It is also important that the format of the effect size is suited for the meta-analytic technique we want to apply and does not lead to errors or biases in our estimate.

• Interpretable: The type of effect size we choose should be appropriate to answer our research question. For example, if we are interested in the strength of an association between two continuous variables, "it is conventional to use correlations to express the size of the effect. It is relatively straightforward to interpret the magnitude of a correlation, and many researchers can understand them..." (Harrer et al., 2022, pp. 53-54).

In summary, the effect size measure needs to be expressed as the same type of measure, only then is it possible to compare them, which of course is the purpose of a meta-analysis. Table 3 depicts forms or types of possible effect size metrics that could be used in a meta-analysis (Higgins et al., 2019; Lipsey & Wilson, 2001).

Table 3

Forms of Effect Sizes (ES) Used Commonly in a Meta-Analysis (Continuous Data)

Type of ES	ES Formula	Formula for Standard Error
Logit-transformed proportions It is used if studies report proportions (p) that is very low or high that leads to a skewed distribution	$p_{logit} = \log_e(\frac{p}{1-p})$	$SE_{P_{logit}} = \sqrt{\frac{1}{np} + \frac{1}{n(1-p)}}$
Pearson Product Moment Correlation It is used to examine the relationship between two continuous variables.	$r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$	$SE_{r_{xy}} = \frac{1 - r_{xy}^2}{\sqrt{n-2}}$
Point Biserial Correlation It is used to examine the relationship between two variables one of which is continues where other is dichotomous.	$r_{pb} = \frac{(\overline{y_1} - \overline{y_2})\sqrt{p_1(1-p_1)}}{s_y}$	
Between Group Mean Difference It is used when there are two independent groups. It is not standardized so, it can be only compared with the studies that measured the outcome using the same scale.	$MD_{between} = \overline{x_1} - \overline{x_2}$	$SE_{MDbetween} = s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where, $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$
Between Group Standardized Mean Difference (Cohen's d) It is used to look the differences in two independent groups. It is standardized so it can be compared with studies measured the outcome using different scales.	$SMD_{between} = \frac{\overline{x_1} - \overline{x_2}}{s_{pooled}}$	$SE_{SMD_{between}} = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{SMD_{between}^2}{2(n_1+n_2)}}$
Within Group Standardized Mean Difference It is used when there is pre and post results for one group.	$SMD_{within} = \frac{\overline{x_{t_2}} - \overline{x_{t_1}}}{\frac{\sqrt{(s_{t_1}^2 + s_{t_2}^2)/2}}{\sqrt{2(1-r_{t_1 t_2})}}}$	$SE_{MD_{within}} = \sqrt{\frac{s_{t_1}^2 + s_{t_2}^2 - (2r_{t_1 t_2} s_{t_1} s_{t_2})}{n}}$

Meta-analytical Models

The end goal of meta-analysis is to establish **one** numerical value that explains the studies as a whole, despite of the fact that the observed effect sizes is actually different for each study. Statistics provides two models, the fixed-effect model and the random-effects model to accomplish the overall goal. Therefore, we are required to make a decision at the beginning of our study. Will it be a fixed-effect model or a random-effects model? Luckily the meta-analytic **pooling** determines the model to be used. In statistics, "pooling" refers to the practice of colleting numerous sets of data that are theorized to have the same attribute (e.g., a mean). These sets of data then are combined to form the "pool". The data in the pool is afterwards used to compute a more precise estimate of that attribute.

The Fixed-effect Model

If all "effect sizes" are assumed to be drawn from a single, homogeneous population then we define this model as the fixed-effect model. This assumption essentially states that all studies used to create the data pool share the same "true effect size". Since all studies are estimators of the same "true effect size" the observed difference in effect size from the "true effect" are due to sampling error. In other words, sampling error causes the observed effect to deviate from the "true effect". Naturally, the larger the sample size the closer or more precise would we in computing the "true effect size".

"The idea behind the fixed-effect model is that observed effect sizes may vary from study to study, but this is only because of the sampling error. In reality, their true effect sizes are all the same: they are fixed (Harrer et al., 2022, p. 95)."

As a consequence, effect sizes reported with relatively smaller standard errors are sampled to compute the pooled effect size. Because of the above reasons we

clearly see that the fixed-effect model is too simplistic in many real-world applications. Therefore, this method is not preferred (Harrer et al., 2022, p. 95).

The Random-effects Model

The random-effects model reflects the reality behind the set of sample data much more realistically. The main assumption is that besides sampling error, there is an additional source of variance. Each study represented in the pool is regarded as an independent draw from a "universe" of populations.

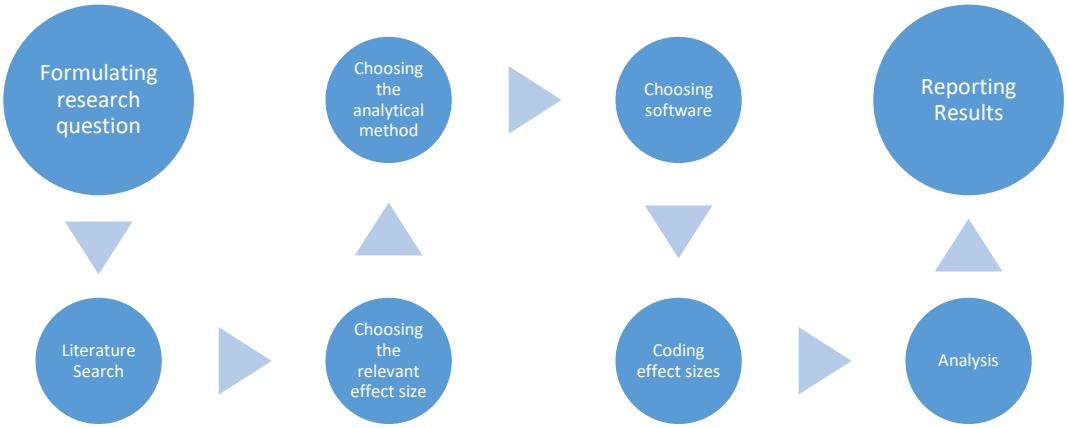
"The random-effects model assumes that there is not only one true effect size but a distribution of true effect sizes. The goal of the random-effects model is therefore not to estimate the one true effect size of all studies, but the mean of the distribution of true effects synthesize the effects of many different studies into one single number (Harrer et al., 2022, p. 100)." Therefore, this method is preferred.

Conducting Meta-analysis

This section presents an eight steps procedure for conducting a meta-analysis that has been used as a norm (Hansen et al., 2021). Furthermore, it provides a theoretical base for our research and serves as a practical guide toward the completion and accountability of this empirical study. Figure 2 below shows at one glance all the sequential steps to be followed in a meta-analysis.

Figure 2

Workflow of Meta-analysis



Step 1: Formulating the Research Question

As with any empirical research, the first item is to formulate the research question. A well-defined and clearly stated research question is needed to form the bases and purpose of the research. The research question constrains the domain of the study. It provides a scope that outlines numerous attributes, such as the target population, variables to be studied, and the problem the study addresses. Finally, it also helps to narrow down and determine the research methodology, sample size, data collection, and data analysis.

This step also puts a natural constrain or limit to the meta-analysis. In order to determine the research questions, there are some suggested procedures to be followed. To name a few:

- Population, Intervention, Comparison, Outcomes (PICO),
- Setting, Population, Intervention, Comparison, and Evaluation (SPICE),

- Sample, Phenomenon of Interest, Design, Evaluation, Research Type (SPIDER), and
- Expectation, Client group, Location, Impact, Professionals, Service (ECLIPSE).

What all procedures have in common presented above is the determination of the population. Since meta-analysis is statistical in nature, it needs the definition of a population. A population is the representation of a distinct group to be studied. In statistics, a population is the pool of the uniqueness from which a statistical sample is drawn for a study. Hence, the determination of uniqueness grouped by common attributes can be said to be a population. We can clearly interfere that SPIDER suggests the population despite not explicitly stated, because it refers to the sample. Or ECLIPSE for example, uses the term “client group” referring to the attributes of the population.

Step 2: Literature Search

To be able to conduct a meta-analysis, the researcher needs to create datasets to be used in the statistical analyses. The dataset is created from prior established or published research. In order to collect a meaningful dataset, and account for reliability it is strongly suggested to conduct the analysis through well-recognized methods. The literature search is actually one of the most important and time-consuming steps in any meta-analysis. The researcher on one hand is competing against time while on the other hand trying to determine the relevant literature through thorough database searches. Luckily, the modern days of electronic databases give now the ability to conduct online database queries. However, an electronic database search brings its own challenges by overloading the collection of queries. Using common sense, filtering procedures based on publication dates,

journal quality, and Boolean operators are all possible tools to limit and improve the literature research. Lastly, it is good advice to include possible old-fashioned hands-on search" for example by utilizing non-academic publications such as magazines and perhaps trying to review relevant unpublished studies.

Step 3: Choosing of Effect Size Measure

The previous section provided information and an extensive discussion on the possible effect size types available for a meta-analysis (see Table 3). Standardized mean differences and correlation coefficient effect size measures are the two most commonly used ones. Needless to say, the researcher is typically constrained by the research question and research field conventions.

Step 4: Choosing the Analytical Method

There are four meta-analytical methods used in education. These are

1. univariate meta-analysis,
2. meta-regression analysis,
3. meta-analytic structural equation modeling (MASEM), and
4. qualitative meta-analysis.

The first three are quantitative, the latter summarizes qualitative findings.

Step 5: Choosing Software

Several software programs are available to compute the effect size metrics of a meta-analysis. The most frequently used software programs are Statistical Package for Social Sciences (SPSS), Stata, CMA, MIX, RevMan, and R. The researchers can decide which software to use considering the time, licensing fees, learning curve associated with the software, amount of the data, and support availability.

Step 6: Creating the Dataset

This step fulfills two requirements simultaneously. The first one can be considered as the template for the digital dataset to be used in the final statistical analysis. The second one is the coding of the variables to be extracted from the journals that eventually will serve as the primary data source. For example, when conducting the univariate method, the coding dataset would be a two-dimensional matrix containing a column of the variable identifiers, the effect size data from the journals as well as the sample size, and each singular journal would be represented by a row in this matrix. One final suggestion is to create initially a large number of columns (representing possible variables) thereby representing the broadest possible depiction of a study. It is always possible to filter the dataset through iterative processes later on. This iterative procedure is especially advisable for MASEM types of methods.

Step 7: Statistical Computation

A robust statistical analysis needs a step concerned with the dataset integrity. The two procedures to do so are detecting possible outliers and influential cases and examining potential publication bias. These are also known as sensitivity analyses.

The first process involves the screening of the possible outlier(s) and influential cases by applying relevant procedures to identify extreme values that deal with this particular concern. However, outliers are values at the extreme ends of a dataset and should not be automatically assumed to be incorrect data. As a result, one has to be careful in the dataset cleansing procedure. In a meta-analysis, if the number effect sizes are small then identifying the outliers becomes particularly important. Outliers can indicate invalid research or simply typing errors when coding the dataset.

Secondly, problems may result in publication bias. For example, studies with interesting or statistically significant findings tend to be published more. So, studies with small samples with unwanted results may not be included in the meta-analysis dataset. There are several methods, such as the funnel plot, Egger's regression test, Peter's regression test, and Duval and Tweedie's trim and fill method, can be used to detect possible publication bias.

Later, one of the two meta-analytical models, fixed or random, is chosen. According to the different statistical assumptions of these models, we justify the heterogeneity between the articles differently. The fixed-effect model assumes that the heterogeneity results from sampling bias, whereas the random-effects model assumes different sources of variance (as mentioned in pp. 43-44).

Step 8: Reporting Results

This step holds true for any statistical report. The report of a meta-analysis should include all crucial output information statistics, such as the mean effect sizes, standard errors, confidence intervals, the number of observations, the study samples, and possible outliers and influential cases. Also, it is important to emphasize the probability value used to reject null-hypothesis. Although in traditional statistics it is common to use the 0.05 value, in meta-analyses experts recommend 0.10 as the statistical heterogeneity have low power (Petitti, 2001, p. 3628). A forest plot is accompanying the tabular output is most often recommended because it provides a visual representation and "a good overview of the different findings and their accuracy" (Hansen et al. 2022 p.12).

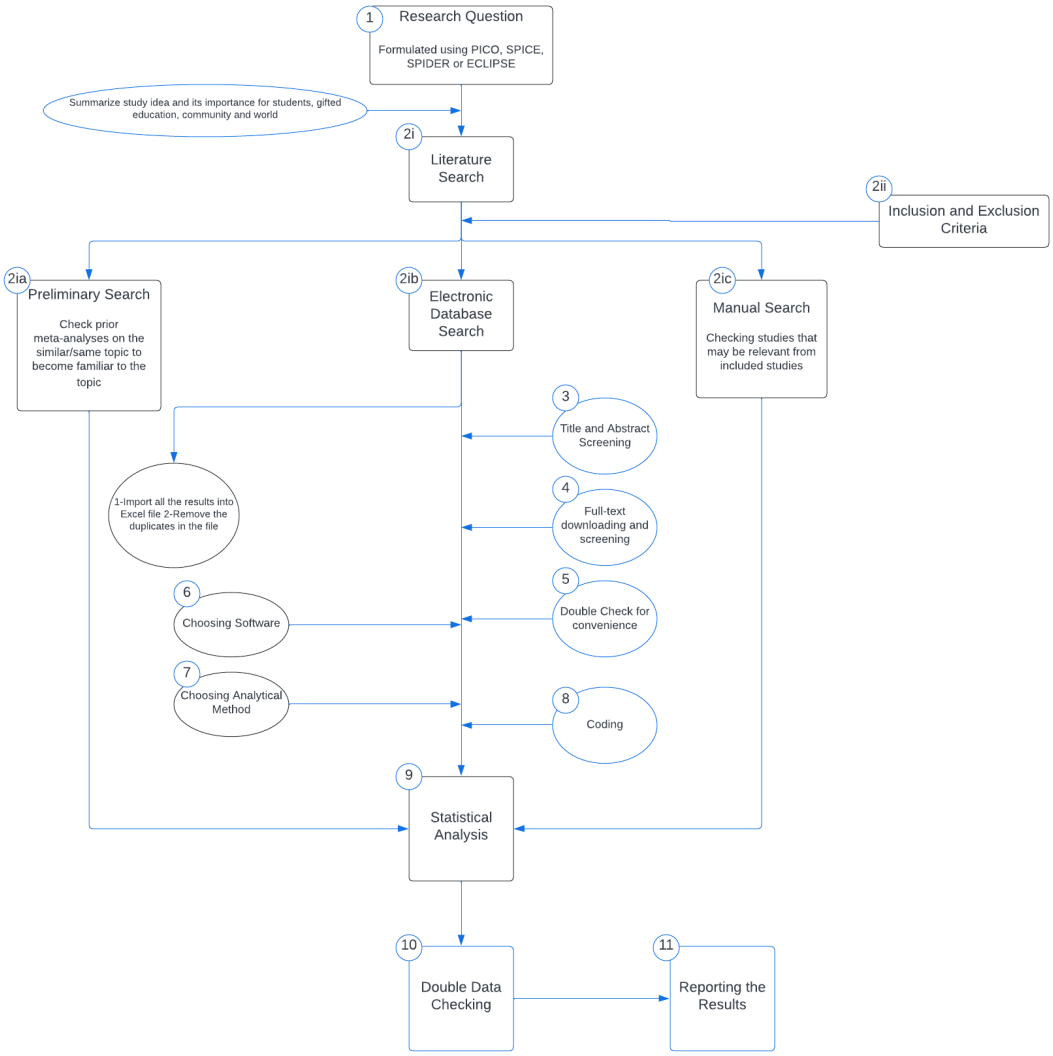
Meta-analysis Conducted

This section "meta-analysis conducted" provides an account on the actual steps used in the study. Obviously, the previous section served as template that was

followed. The flowchart diagram shown in Figure 3 is a pictorial reference depicting the steps followed.

Figure 3

Flowchart of the Meta-analysis Conducted



Research Question

The purpose of this meta-analysis was to reach an overall understanding of the effect of the enrichment programs on the academic achievement of gifted and talented students. Hence the research question was determined to be: how do

enrichment programs affect the academic achievement of gifted and talented students?

The PICO framework (see p.45), a common framework that aims to define the statistical variables was involved in a meta-analysis was used (Eriksen & Frandsen, 2018). Table 4 shows the main four variables and its definition in this research.

Table 4

PICO for This Thesis

PICO Abbreviation	Meaning in this thesis
Population	Gifted and talented students
Intervention	Enrichment
Comparison	No enrichment
Outcome	Academic achievement

Literature Search

Education Resources Information Center (ERIC), Bilkent online library, Google Scholar and YÖKTEZ were used to find relevant studies. ERIC is a free database that contains around two million information on research and publications in the education field. It includes journal articles, research reports, fact sheets, conference papers, books, and non-journal articles' information and if possible, the actual documents as well. So, the first attempts to locate the articles were performed using ERIC. Bilkent library online was used to reach the actual articles. After that, Google Scholar search was conducted as well. Google Scholar search was conducted to minimize the file drawer problem (Gusenbauer & Haddaway, 2019). I also used Yükseköğretim Kurulu Tez Merkezi (YÖKTEZ). YÖKTEZ database involved the entire thesis done in Turkey. There were no new studies from that database. The search results were restricted to the studies published between the years 2013 and 2021.

The following keywords and their combinations were used to find the studies: *gifted students, talented students, enrichment programs, academic achievement, pull-out programs,* and *high ability*.

All the search results were initially documented as a spreadsheet file. Then, the researcher checked the document to determine the duplicates using Excel's conditional formatting feature. The electronic database search of the keywords matching the search criteria resulted in a total of 4659 records. Then removal procedure of duplicates was applied which resulted in 3505 unique articles. The tedious and time-consuming procedure of "Titles and abstracts" screening for appropriateness for the inclusion criteria (see next section) resulted in the removal of 3360 studies. The full texts of the remaining 145 studies were then obtained and read for eligibility. Also, the reading included a cross checking of references (i.e. citations) that were used these articles. Among the 145 articles, 123 articles were removed. The main reasons for these full-text exclusions were the use of tasks that were outside the stated scope of the review; performance measures were self-reports of satisfaction with performance; there was no quantitative assessment of the flow–performance relationship. This resulted in a final sample of 22 relevant articles. A list of the keywords and associated numbers of the articles are listed in Table 5 below. In the first column one can see the search engine used for the search. Next to it, one can see the keywords that were searched. All these searches were limited to the dates 2013 and 2021. In the third column, one can see the number of results for each keyword set. All the results were checked with the previous search results to find duplicates. The numbers of these unique results were reported in the fourth column. All these results were scanned by reading their titles and abstracts. After some of the studies were excluded, possible studies that might be included were read

in detail. The number of these results was given in the fifth column. Lastly, the numbers of studies that are included in the study were shown.

Table 5

Number of Search Results

Search Engine	Keywords	All results	Unique results	Articles read full text	Included studies
ERIC	“enrichment programs” achievement	232	232	25	2
ERIC	“enrichment” “achievement”	323	91	10	0
ERIC	“gifted” “enrichment”	139	96	28	1
ERIC	Enrichment programs for gifted	951	735	19	5
ERIC	"gifted students" "enrichment program" "academic achievement"	8	0	0	0
ERIC	"gifted students" "enrichment program" "academic"	13	0	0	0
ERIC	"high-ability" "enrichment"	20	2	0	0
ERIC	"pullout programs"	14	8	1	0
Google Scholar	“üstün yetenekli” “zenginleştirme”	590	539	28	8
Google Scholar	"gifted students" "enrichment program" "achievement" "experiment"	276	264	15	2
Google Scholar	"gifted students" "enrichment" "mathematics achievement"	715	623	9	2
Google Scholar	"gifted students" "enrichment" "science achievement"	391	229	3	1
Google Scholar	"gifted students" "enrichment" "language achievement"	53	30	3	0
Google Scholar	"gifted students" "enrichment" "biology achievement"	11	3	0	0
Google Scholar	"gifted students" "enrichment" "chemistry achievement"	14	3	0	0
Google Scholar	"gifted students" "enrichment" "physics achievement"	15	3	0	0
Google Scholar	"gifted students" "enrichment" "history achievement"	2	1	0	0
Google Scholar	"gifted students" "enrichment" "geography achievement"	1	1	0	0
Google Scholar	"gifted students" "enrichment" "gpa"	891	648	7	0

Inclusion and Exclusion Criteria

After the removal of all duplicates, the researcher read all the abstracts. In the cases where there was doubt, other parts of the studies (especially the method) were read to decide whether it was appropriate to include or exclude the article/study. All the potential studies were kept in an excel file and discussed with the advisor later.

Inclusion criteria:

1. The medium language of the studies was either English or Turkish
2. The studies had information to calculate the effect size.
3. The publication date was between the years of 2013 and 2021.
4. The investigation of effectiveness of an enrichment program on academic achievement of gifted and talented students.
5. The sample consists of students from grades K12.
6. The research design was quantitative.

Exclusion criteria:

1. The articles have sampled other than gifted and talented students only.
2. The articles that have samples twice are exceptional.
3. Research focused on other issues than the effects of an enrichment program, such as students' identification process, parents' involvement, and teachers' professional development.
4. Studies focus on variables like creativity, leadership, and interest in academic areas. However, these were included if they were explicitly stated or defined as academic achievement by the authors.

Electronic Database Search

All articles retrieved from the searches were initially read and checked for eligibility against the eligibility checklist described above. Then, 52 articles were

discussed with the advisor. This process was performed independently. First, both the author and the advisor read the articles, and then they discussed and made the decision for either excluding or including the articles.

Effect Size Measure Choice

Effect size is a quantitative value representing the strength and the direction of the relationship between two variables in a population or a sample-based estimate of that quantity. This research shows the relationship between enrichment programs and gifted and talented students' academic achievement.

The type of the effect size used in this research is between group standardized mean difference (SMD), also known as Cohen's d, as it is used to observe the differences in two independent groups. The two independent groups in this study are the experimental and control groups. Also, SMD is standardized, which means that it can be compared with studies that measure the outcome using different scales. In this meta-analysis, different scales were included in the included studies as they measure academic achievement differently (see Table 3.) After that, the d value is converted into Hedges' g to minimize the small sample bias (Harrer et al., 2022, p.81).

Meta-analytical Model Choice

In this study, a univariate meta-analysis model was selected. The reason to choose this model was that our research question focuses on the relationship between the enrichment programs and students' academic achievement. To do so, an effect size was calculated using statistical analysis in a quantitative way. However, this model has a disadvantage as it is weak in terms of explaining the observed variance among findings (Gonzalez-Mulé & Aguinis, 2018 as cited in Hansen et al., 2021). Therefore, seven hypotheses were tested to identify possible moderators.

Table 6

Included Studies

Study	Date	Ss #	Grade	Language	Research Type	Pub. Type	Identical Pre-Post
Al-Zoubi	2014	n:30	6-8	English	Pre-Post	Journal	Yes
Altıntaş	2015	n_e:15 n_c: 15	6-8	English	Exp-Cont	Journal	No
Altıntaş	2014	n_e:6 n_c: 6	K-5	Turkish	Exp-Cont	Doctoral Dissertation	No
Baser	2016	n:5	K-5	English	Pre-Post	Doctoral Dissertation	No
Booij	2017	n_e:1067 n_c:1356	9-12	English	Exp-Cont	Non-Journal	No
Çalıkoğlu	2014	n_e:10 n_c: 9	K-5	Turkish	Exp-Cont	Doctoral Dissertation	Yes
Economos	2021	n_e:32 n_c: 39	9-12	English	Exp-Cont	Doctoral Dissertation	No
Erdoğan	2014	n_e:11 n_c: 10	6-8	Turkish	Exp-Cont	Doctoral Dissertation	Yes
García-Perales	2019	n_e:9 n_c: 9	K-5	English	Exp-Cont	Journal	Yes
Karataş	2013	n_e:12 n_c: 12	K-5	Turkish	Exp-Cont	Doctoral Dissertation	Yes
Koehler	2013	n_e:7 n_c: 7	9-12	English	Exp-Cont	Doctoral Dissertation	No
Korkut	2017	n_e:12 n_c: 12	K-5	Turkish	Exp-Cont	Master's Thesis	Yes
Kuykendall	2020	n_e:69956 n_c:23935	K-5	English	Exp-Cont	Doctoral Dissertation	No
Laughlin	2013	n_e:23 n_c:23	K-5	English	Exp-Cont	Doctoral Dissertation	No
Mccoach	2014	n_e:205 n_c:127	K-5	English	Exp-Cont	Journal	No
Olszewski-Kubilius	2017	n_e:204 n_c: 744	K-5	English	Exp-Cont	Journal	No
Özçelik	2017	n_e:14 n_c: 14	K-5	Turkish	Exp-Cont	Doctoral Dissertation	Yes
Preckel	2019	n_e:137 n_c: 128	6-8	English	Exp-Cont	Journal	No
Üşenti	2013	n_e:11 n_c: 11	K-5	Turkish	Exp-Cont	Doctoral Dissertation	Yes
Van der Muelen	2013	n:26	K-5	English	Pre-Post	Journal	Yes
Yu	2020	n:28	K-5	English	Pre-Post	Journal	Yes
Zayac	2013	n_e:563 n_c: 257	6-8	English	Exp-Cont	Doctoral Dissertation	No

Table 6 is the resulting dataset and provides an overview of the study characteristics (participants, publishing date, sample size, grade level, language,

design, publication type, and whether they used the identical pre-test or post-test in the study). The sample consisted of 22 articles, all published between the years of 2013 and 2021, with almost half of the being published after 2016. Of the 22 studies, 16 have sample size less than 100.

Software Selection

There were two main reasons why the researcher chose R as the software program to conduct the statistical analysis of this meta-analysis. R is a very advanced statistical software application that has proven itself and used widely by professionals from a divers of research background. This program is free of charge. The other reason was that R allows future researchers to reproduce the analysis conducted here. All the codes, algorithms and packages are listed in the appendix to encourage in replication of our results. A new community of researchers conducting meta-analysis, actually require the dataset for reliability and validity purpose (see Appendix B).

Data Extraction

Before running the actual statistics, the final step included the data extraction from the 22 studies determined. All studies were read at least twice to make sure the coded data were correct (Cuijpers, 2016; Petticrew & Roberts, 2008). Based on recommendations by Popay et al. (2006), the following data fields were extracted: authors, date of publication, language of the study, sample size, grade level, design, and whether their pre-test and post-tests are identical (see Appendix A).

Treatment of Multiple Effect Sizes in a Study

Only one effect size from each study was included in the meta-analysis to prevent overrepresentation of any study and violation of the independency assumption of the studies. To do so, if the articles provided total achievement data,

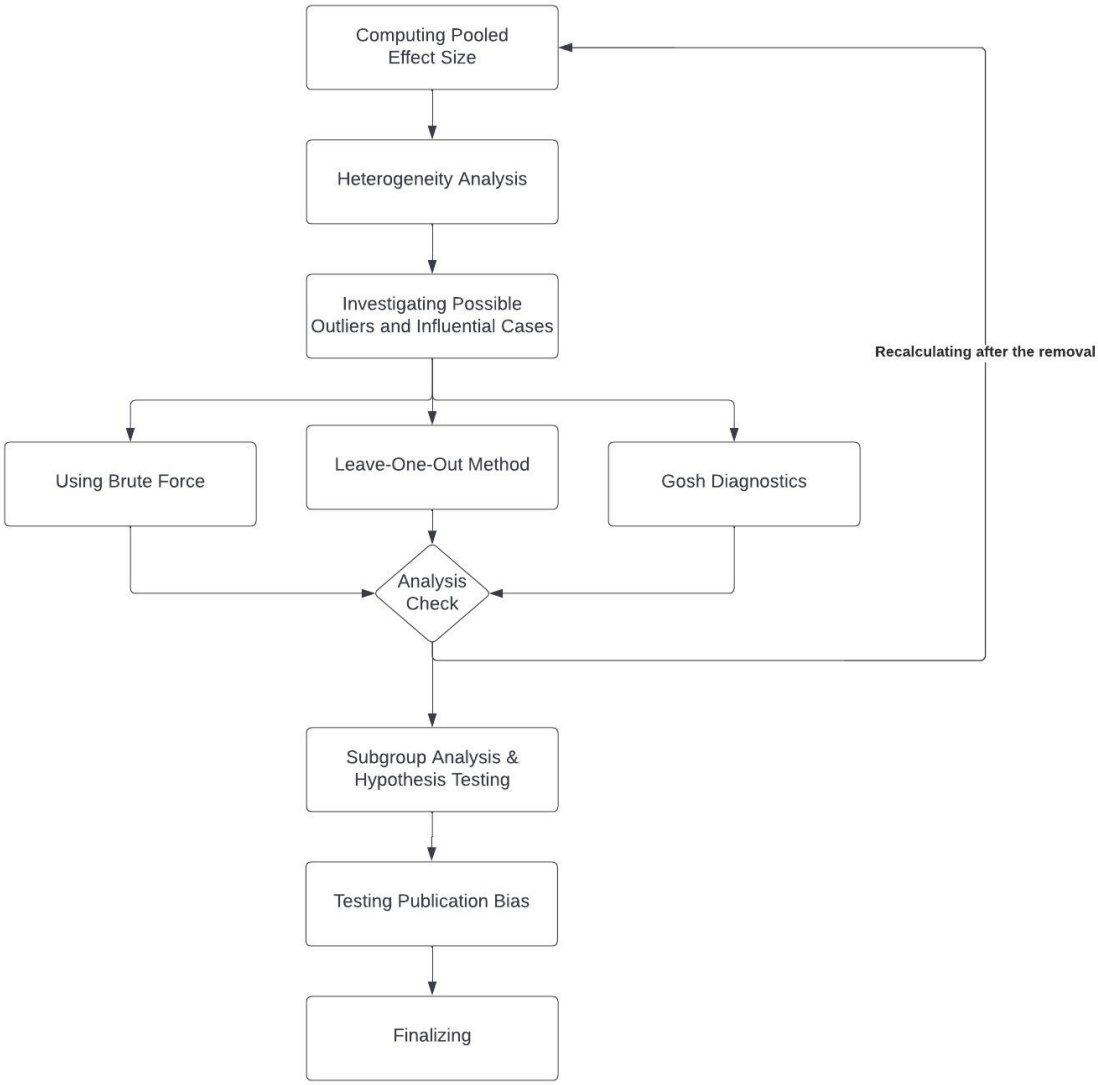
then only this data was used. Otherwise, each effect size was given a number, and the one that will be included was selected using Google's random number generator.

Analysis

In order to answer the research question and report on the robustness of it, the following statistical procedure shown in Figure 4 was followed.

Figure 4

Steps Followed in Statistical Analysis



As shown in Figure 4, the initial step was to compute the overall effect size. Then the other steps were used to test the rigidity (robustness degree) of the overall effect size. This was done through two main processes: heterogeneity analysis and testing for publication bias. To investigate the between-study heterogeneity, possible outliers and influential cases detected through brute force, leave-one-out, and Graphic Display of Heterogeneity (GOSH) methods. After that, the possible reasons for heterogeneity were examined by performing subgroup analyses. So far, these procedures allow us to pool the effect size and investigate heterogeneity along with its possible reasons. The final step is testing publication bias.

CHAPTER 4: RESULTS

Introduction

This chapter discussed the analyses conducted in line with the research question were discussed. In this study, the following research question was examined:

Research Question: How do enrichment programs affect the academic achievement of gifted and talented students? To report on the degree of robustness of the computed overall effect size (OES), the following hypotheses posing potential heterogeneity were explored.

Null hypothesis 1: There is no difference between the effect size of the studies that have a small sample size (n<100) and a large sample size (n>100).

Null hypothesis 2: There is no difference between the study's effect sizes with different grade levels of students as their sample.

Null hypothesis 3: There is no difference between the effect size of the studies that are written in English and Turkish.

Null hypothesis 4: There is no difference between the effect size of the studies that are designed as one group pre-post test and control treatment groups.

Null hypothesis 5: There is no difference between the effect size of the studies that are published in the journal and that are doctoral dissertations that are not published.

Null hypothesis 6: There is no difference between the effect sizes of the studies that are using the same pre-test and post-test and different pre-tests and post-tests.

Null hypothesis 7: There is no difference between the effect sizes of the studies that are investigating different subject areas.

Results of the Study

Research Question: How do enrichment programs affect the academic achievement of gifted and talented students?

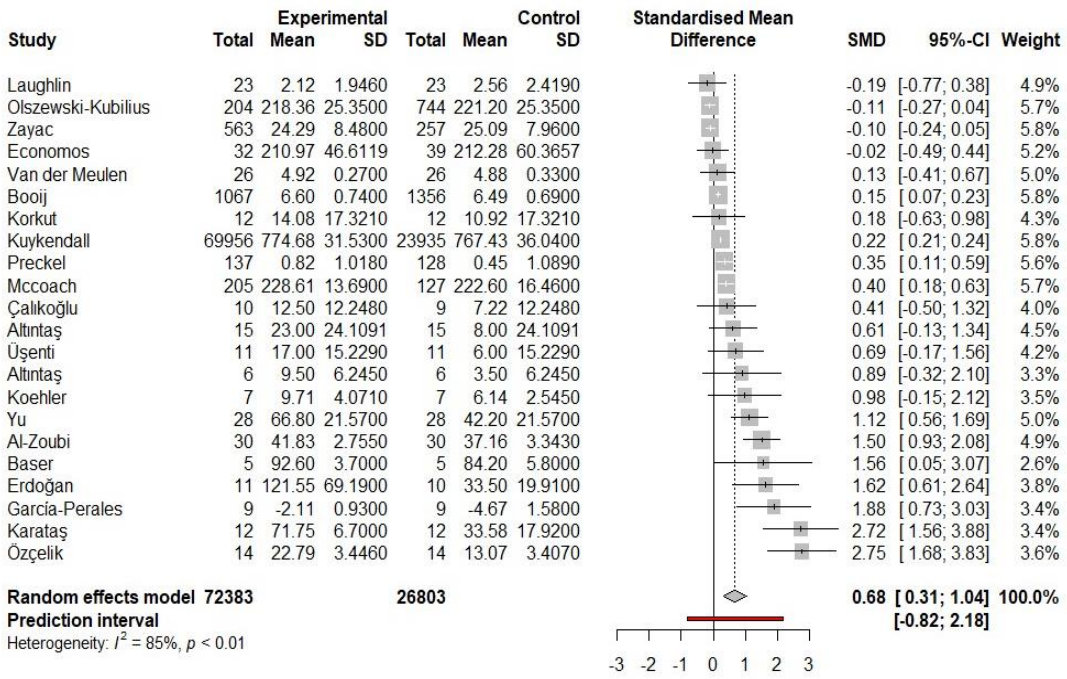
The results of the steps followed (see Figure 4) will be represented to answer this research question.

Computing the Overall Effect Size

In this study, the standardized mean differences under the random-effects model for each study were calculated. There were 22 studies included in this meta-analysis. The total number of observations was 99186. Figure 5 below is a forest plot (also called Blobbogram) developed specifically for reporting meta-analysis' findings. A forest plot provides a visual representation of statistical estimates that compares the scientific studies researching the same phenomenon. The visual representations are similar to ordinary box plot graphs. The gray boxes represent the effect size value (SMD g) of the individual studies, and the associated horizontal lines indicate the confidence intervals.

Figure 5

The Forest Plot for Included Studies



In this analysis, inverse variance method was used. This meant that the weight of each study was inversely proportional to its variance. Low variance indicates high precision. The dimension of the grey boxes reflects the weight of each study. More specifically, the first four studies had negative effect size values. This fact is also seen on the left side of the chart, where the statistics of the experimental and control group were given. The Figure 5 shows the majority of interventions (18 out of 22) were in favor of the experimental group.

At the very bottom of the plot, the overall effect size (SMD g =0.68) with the 95% confidence interval ranging from 0.31 to 1.05 is denoted by a diamond shape. The thick line under the diamond shape represents the prediction interval. The effect size was positive, which means that the academic achievement of the treatment group was higher than the control group. An SMD g value of about 0.50 denotes a medium effect (Harrer et al., 2022, p.66). So, the findings can be interpreted as a

medium effect. Since p=0.001, which is less than 0.10, the effect was statistically significant.

There is a dotted line perpendicular to the center of the diamond, and one can see that the studies that have a smaller effect size than the pooled effect size are on the left of this line, whereas studies that have a larger effect size are on the right of this line. One can see that the effect sizes of the last two studies, namely Karataş's (SMD g =2.72) and Özçelik's (SMD g =2.75), are quite far from this line which was an indication of possible outliers. It is noteworthy that the interval contained negative values indicating that future studies could yield negative results (as mentioned in the previous section).

Another finding worth mentioning was that there are positive and negative SMD (g) effect size values. This indicated that treatment output or interventions conducted on gifted students yielded both positive and negative effect on students' achievement. It was worth noting that the two highest SMD (g) values belong to Özçelik (2017) (SMD g = 2.75) and Karataş (2013) (SMD g =2.72), both Turkish works. The lowest was Laughlin's, which was -0.20.

Moreover, from the graph portion of Figure 5, we observed that the boxes were comparable with respect to their dimensions; that is, none of the boxes were too big or small compared to each other. However, we also observed six cases in which the tails were inside the boxes. These were Booij (2017), Kuykendall (2020), Olszewski-Kubilius (2017), Mccoach (2014), Preckel (2019), and Zayac (2013). These studies have larger sample sizes than the rest of the studies, making them more precise. A careful observer may notice that the line in Kuykendall's study is so small that it is almost impossible to see. However, this is also an indication of possible an *influential* case.

Heterogeneity Analysis

The second step followed in the statistical procedure of this meta-analysis was to report on heterogeneity, that is, the robustness of the overall effect size. A high to very high heterogeneity may mean that we might be comparing two unrelated factors and hence, indicating that the studies in these meta-analyses had nothing in common. This phenomenon is also called the apples and oranges problem. On the other hand, a low heterogeneity value may lead the researcher to revert to the fixed-effects model analysis. The fixed-effects model assumes the only error results from errors within the studies (see chapter 3). As stated in chapter 3, the random-effects model computes an overall effect size at all costs, so reporting only the overall effect size limits the interpretation (Harrer et al., 2022, p.139). Heterogeneity analyses may be considered a follow-up statistic, further exploring the meaning of the computed overall effect size (Baujat et al., 2002, p. 2651). Therefore, one needs to report heterogeneity in conjunction with the overall effect size. The three statistical heterogeneity measures are, τ^2(tau square), I^2 (I square), and Cochran's Q. Table 7 shows the result of the heterogeneity τ^2, and I^2 results.

Table 7

$\tau^2$and I^2 Heterogeneity Analyses

τ^2	τ	I^2	Prediction Interval
0.49	0.70	84.6%	
[0.25, 1.34]	[0.50, 1.15]	[77.9%, 89.3%]	[-0.82, 2.18]

The prediction interval ranged from -0.82 to 2.18, indicating that negative intervention effects cannot be ruled out for future studies (Harrer et al., 2022, p.152). An I^2=84.6% refers to substantial heterogeneity (Harrer et al., 2022, p.146). Also, the confidence interval of τ^2 does not contain zero, which is another indication that

possible between-study heterogeneity existed (Harrer et al., 2022, p.151). The value τ is the square root and is provided in the Table 7 as an old convention (Harrer et al., 2022, p.147).

Table 8

Cochran's Q Heterogeneity Measure

Q	*df*	*p*
136.44	21	<.001

Table 8 shows that the third heterogeneity measure, namely the Cochran's Q value. The p value was statistically significant hence confirming heterogeneity. This finding was in agreement with $\tau^2$and I^2 results provided Table 7 above. Since high heterogeneity was established we need to analyze the data further.

Investigating Possible Outliers and Influential Cases

The results reported so far were based on the statistics of all the 22 studies that were included in this meta-analysis. However, as pointed out in the forest plot (see Figure 5), the results indicated possible outliers and/or influential cases which may have had a significant impact on the overall effect size (SMD g). There are three ways to detect these studies (see Figure 8): Brute force, leave one out, and GOSH plot method.

Brute Force Method

The first way is to determine whether the confidence interval of a single study overlaps with the overall effect size's confidence interval. If it does not overlap, then it is called a potential outlier. In this study, the studies conducted by; Booij (2017), Kuykendall (2020), Olszewski-Kubilius (2017), Mccoach (2014), Preckel (2019), and Zayac (2013) are potential outliers under the random-effects model as their confidence interval do not overlap with the overall confidence

interval. For instance, Olszewski-Kubilius' 95% confidence interval was [-0.27, 0.04], where the pooled confidence interval for this study is [0.31, 1.04]. In order to find out if these studies (Booij (2017), Kuykendall (2020), Olszewski-Kubilius (2017), Mccoach (2014), Preckel (2019), and Zayac (2013)) impact the overall effect size, *they were removed from the dataset and the overall effect size was recomputed again*. Table 9 shows the results of the recomputed overall effect size after removing the studies mentioned above.

Table 9

Pooled Effect Size of Studies after Removing Outliers (Random Effects Model)

SMD (g)	%95 CI	*p*
0.64	[0.31, 0.97]	<.001

After the removal, SMD g dropped from 0.68 to 0.64, which is again a medium effect in favor of the treatment group (Harrer et al., 2022, p.66). So, removing the potential outliers and influential cases detected by brute force did not have a major effect on the value of the overall effect size. The confidence interval became smaller, which means the result was more precise, and the p-value is still statistically significant, meaning heterogeneity existed. However, it is important to recheck the heterogeneity value as previously, the heterogeneity was large. Tables 10 and 11 show the heterogeneity analyses of the results, when outliers were removed.

Table 10

Heterogeneity Analyses after Removing Outliers

τ^2	τ	I^2	Prediction Interval
0.23	0.48	66.3%	
[0.06, 0.77]	[0.25, 0.88]	[42.9%, 80.1%]	[-0.43, 1.71]

The prediction interval ranged from -0.43 to 1.71, indicating that negative intervention effects cannot be ruled out for future studies (Harrer et al., 2022, p.152). The prediction interval narrowed down. This was expected (Harrer et al., 2022, p.150). TheI^2 value drops to 66.3%. This outcome can be classified as a moderate heterogeneity (Harrer et al., 2022, p.146). This was also expected as the I^2 value is proportional to the sample size. It decreases as the sample size decreases. The confidence interval of τ^2 still does not contain zero, which is another indication that possible between-study heterogeneity existed (Harrer et al., 2022, p.151).

Table 11

Cochran's Q Heterogeneity Measure after Removing Outliers

Q	*d.f.*	*p*
44.47	15	<.001

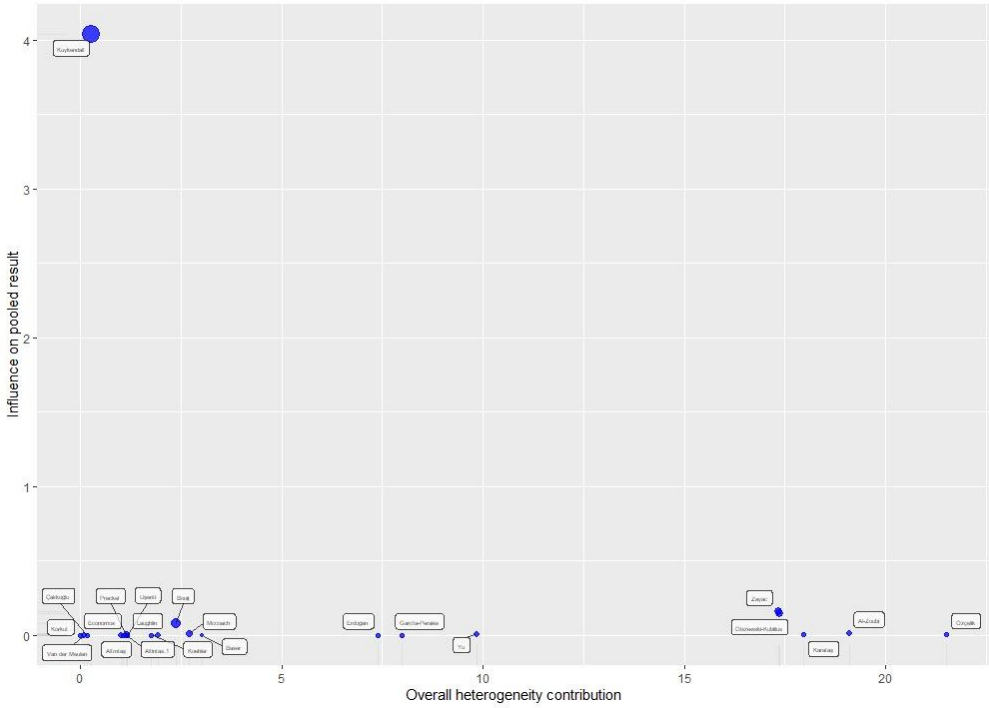
Similarly, the existence of possible heterogeneity still holds by looking at the chi-square output provided in Table 11 with Q (15) =44.47. The results show that the findings become more precise after the removal but do not significantly change the magnitude and direction of the effect size.

Leave-one-out Method

The second method used to determine heterogeneity is known as the "Leave one out" method Baujat et al (2002). In the leave-one-out method, robustness is explored iteratively by excluding each time one of the studies. Because we had 22 studies, 22 iterations were run. Figure 6 below shows the Baujat plot for this study.

Figure 6

Baujat Plot



Baujat plots are used to detect the studies that have a large impact on Cochran's Q heterogeneity of the pooled findings. The x-axis denotes the overall heterogeneity contribution, and the y-axis shows the influence on the pooled results. When moving from left to the right, the relative contribution to heterogeneity increases (Harrer et al., 2022, p.158). So, we can say that the possible *outliers* were placed at the right side of the plot; when moving upwards, the influence on the pooled effect size increases. Therefore, the studies in the upper part of the plot were possibly *influential* cases.

In the right-end corner of the Baujat plot, the studies; Al-Zoubi (2014), Karataş (2013), Olszewski-Kubilius (2017), Özçelik (2017), and Zayac (2013) appear to be possible outliers. These were consistent findings we commented earlier in the above section through an initial visual screening of Figure 5. We also observed that Kuykendall's study shown above is on the left side. This means that it makes

almost no contribution to heterogeneity. Therefore, we can conclude that Kuykendall (2020) can be considered an influential case but not an outlier.

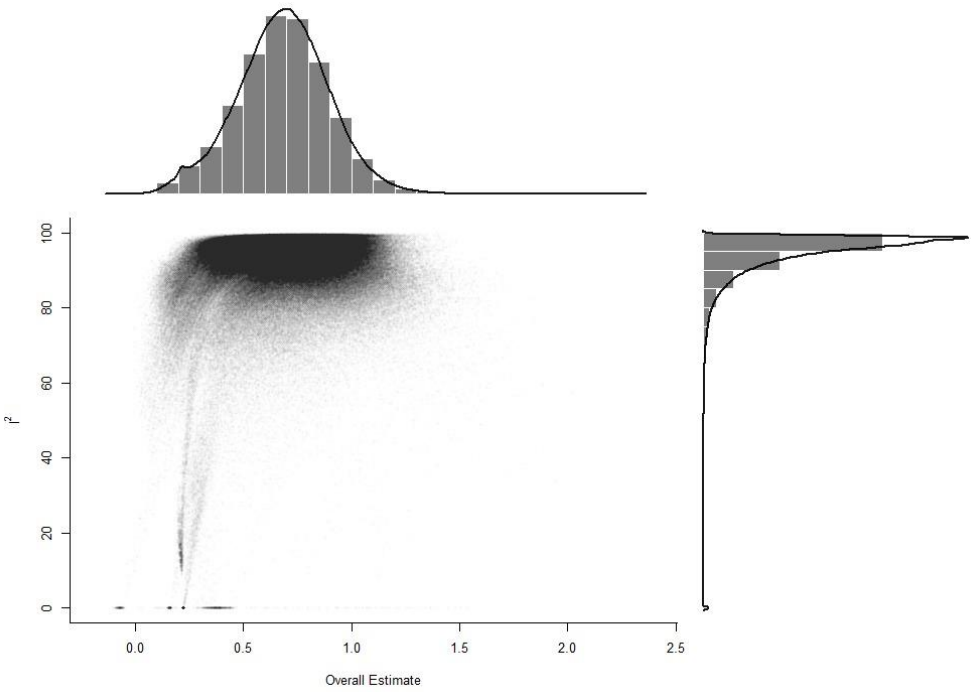
Hence under the externally standardized residuals diagnostics, Karataş (2013) and Özçelik (2017) studies deviate most from the pooled effect size, making them strong candidates for outliers.

Graphic Display of Heterogeneity (GOSH) Plot Analysis

Another method that involves more complex analysis is called GOSH plot analysis. Rather than looking at the findings after the removal of each study one by one, this plot allows the researchers to examine all possible subsets of the included studies. In this study, there were 22 possible combinations in the leave-one-out model, whereas there were 2^{22}-1 (4,194,303) possible combinations for the GOSH model. This makes the GOSH analysis more complex. Each meta-analysis result is shown with a dot in the plot. If the sample of the meta-analysis contains homogeneous studies, then the GOSH plot is roughly symmetric and homogeneous. This means that the effect found in the study is robust. Otherwise, the two or more clusters signal the existence of more than one effect size population in the data. In that case, performing a subgroup analysis may allow the researcher to the reason for these clusters (Harrer et al., 2022, p. 163). Figure 7 shows the GOSH plot for this meta-analysis.

Figure 7

GOSH Plot



The pattern we see in this GOSH plot is that the overall estimate (x-axis) is unimodal; the histogram has only one peak. This outcome also holds true for the I^2 values (y-axis) as well; this distribution is unimodal. Furthermore, the overall estimate size is almost symmetric, where I^2 values are heavily skewed to the right. The black accumulation shows that the majority of the studies have high effects and high heterogeneity. However, there are some combinations where the heterogeneity and effect size are small. Table 12 provides the outcome of our GOSH plot.

Table 12

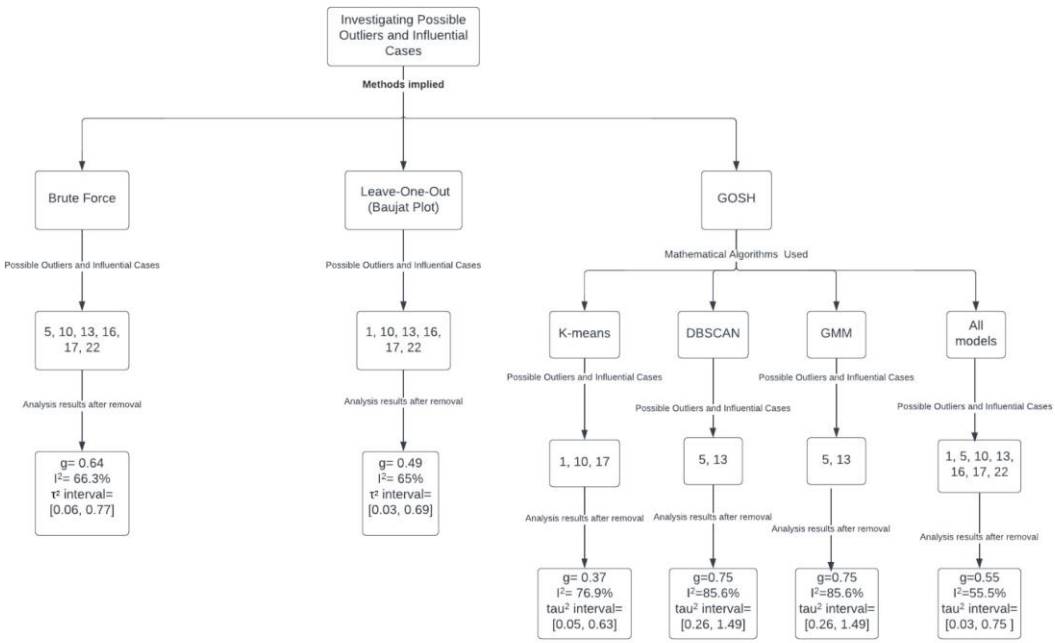
Variables to Check when Exploring the GOSH Plot

Modal	Compactness	I^2
Uni	Compact	High
Multi	Not Compact	Low

In this study, the I^2 values are compact, with the predominance of values on the y-axis between 50% and 100%. This shows high heterogeneity, but there is no sign of any clusters. The majority of the meta-analysis combinations displayed in Figure 7 have an overall estimate between 0.25 and 1.0, with heterogeneity of 80% and above. These results warrant for subgroup analyses.

Figure 8

Summary of the Investigation of Possible Outliers and Influential Cases



Subgroup Analysis and Hypothesis Testing

The analyses so far on heterogeneity gave us indications that some studies might not fit the overall effect without any reasons to why this was the case. Subgroup analyses are used to test possible hypothesis for the reasons. In this research, the subgroup analysis was conducted on the sub-queries posing potential heterogeneity. Particularly the following corresponding null hypotheses were conducted: The number of studies analyzed for each hypothesis was shown in Table 13.

Null hypothesis 1: There is no difference between the effect size of the studies that have a small sample size (n<100) and a large sample size (n>100).

Null hypothesis 2: There is no difference between the study's effect sizes with different grade levels of students as their sample.

Null hypothesis 3: There is no difference between the effect size of the studies that are written in English and Turkish.

Null hypothesis 4: There is no difference between the effect size of the studies that are designed as one group pre-post test and control treatment groups.

Null hypothesis 5: There is no difference between the effect size of the studies that are published in the journal and that are doctoral dissertations that are not published.

Null hypothesis 6: There is no difference between the effect sizes of the studies that are using the same pre-test and post-test and different pre-tests and post-tests.

Null hypothesis 7: There is no difference between the effect sizes of the studies that are investigating different subject areas.

Table 13

Sample Size of the Subgroups

Subgroup Analysis	Variables	k
Sample size	Small	16
	Large	6
Grade level	K-5	14
	6-8	5
	9-12	3
Language	English	15
	Turkish	7
Design	Pre-Post	4
	Exp-Cont	18
Publication Type	Journal	8
	Non-journal	1
	PhD	12
	Master's	1
Tests used	Same Pre-Post	10
	Different Pre-Post	12
Subject area	Total	3
	Mathematics	11
	Science	4
	Language	2
	Social Science	2

Table 14

Subgroup Analysis by Sample Size

Sample Size	k	*g*	*95% CI*	I^2	$p_{subgroup}$
Small	16	0.97	[0.49, 1.45]	78%	<.001
Large	6	0.14	[-0.08, 0.37]	88%	

Table 14 shows the total number of included studies, the number of the studies (k) with small sample size was 16 whereas; the number of the studies with

large sample size was 6. An SMD (g) value of about 0.80 denotes a large effect size. On the other hand, an SMD (g) value of about 0.20 denotes a small effect size (Harrer et al., 2022, p.66). The effect size (g=0.97) of the studies with small sample size is considerably larger than the effect size (g = 0.14) of the studies with a larger sample size (see Table 14). This means that the means of the treatment group changed more in the small sample size studies. However, heterogeneity in terms of I^2 shows a large heterogeneity for both groups. The p-value is less than 0.10, so the results are statistically significant. We reject the Null Hypothesis.

Table 15

Subgroup Analysis by Grade Levels

Grade Level	k	*g*	*95% CI*	I^2	$p_{subgroup}$
K-5	14	0.79	[0.24, 1.33]	85%	.013
6-8	5	0.72	[-0.20, 1.63]	91%	
9-12	3	0.15	[-0.05, 0.35]	24%	

Table 15 shows that most of the studies included were from K-5. Only three studies targeted high-school students. The effect size of the studies with samples consisting of grades K-5 is nearly the same as the grades 6-8 and they are both large whereas, it is small for grades 9-12. Besides, heterogeneity in terms of I^2 does not differ that much for grade levels K-5 and 6-8. The p-value is less than .10, so the results are statistically significant.

Table 16

Subgroup Analysis by the Publication Language

Language	k	*g*	*95% CI*	I^2	$p_{subgroup}$
English	15	0.42	[0.08, 0.75]	84%	.044
Turkish	7	1.28	[0.30, 2.27]	77%	

Table 16 shows the effect sizes of the studies that are written in English can be considered small to moderate. The Turkish ones on the other hand, show a large effect. This is expected as one can see from Figure 5 that only two Turkish studies were in the first half of the studies. In addition to this, the last two studies which show the largest effect size belong to Turkish studies. Heterogeneity in terms of I^2 does not differ as both groups have large heterogeneity. However, one can observe that the confidence interval for Turkish studies is broader which means they are less precise. The p-value is less than .10, so the results are statistically significant. Therefore, the language of the studies may be a reason that contributes to the heterogeneity.

Table 17

Subgroup Analysis by the Study Design

Study design	k	*g*	*95% CI*	I^2	$p_{subgroup}$
Pre-post	4	1.00	[-0.07, 2.07]	77%	.032
Exp-Cont	18	0.61	[0.19, 1.04]	84%	

Table 17 shows there were only four studies that have one group pre-test, post-test design, so the number is small. The effect size of the studies that followed one group pre-test and post-test design had a considerably large effect than the studies that had an experimental-control design. The 95% confidence interval for pre-test post-test design contains both negative and very big positive values. It is interesting as the variance seems high. In addition, heterogeneity in terms of I^2 does not differ as both groups have large heterogeneity. The p-value is less than .10, so the results are statistically significant.

Table 18

Subgroup Analysis of Journal and Dissertations

Publication Type	k	*g*	*95% CI*	I^2	$p_{subgroup}$
Journal	8	0.64	[0.09, 1.18]	88%	.022
Non-Journal	1	0.15	[0.07, 0.24]	-	
Phd Dissertation	12	0.85	[0.21, 1.50]	85%	
Master's Thesis	1	0.18	[-0.63, 0.98]	-	

According to Table 18, most of the studies were either published in a journal or a doctoral dissertation. Master studies and non-journal articles that are not a thesis or a dissertation have only one article, so only the journal articles and doctoral dissertations are compared. The effect size of doctoral dissertations has a considerably larger effect size than dissertations. This is interesting as it is advised to include grey literature suggesting that the published articles have a larger effect size due to the publication bias (McAuley et al., 2000, as cited in Harrer et al., 2022, p.230). That was not the case for this study. Heterogeneity in terms of I^2 does not differ much as both groups have large heterogeneity. The p-value is less than .10, so the results are statistically significant.

Table 19

Subgroup Analysis of Studies by the Tests They Use

Same tests used in pre-test and post-test	k	*g*	*95% CI*	I^2	$p_{subgroup}$
Yes	10	1.23	[0.56, 1.91]	78%	<.001
No	12	0.17	[-0.02, 0.36]	78%	

Table 19 shows the number of studies that used the same test as pre-test and post-test. Others either change the pre-test a little or used a completely different test as a post-test. The effect size of the studies that used the same test for pre-test and

post-test had a considerably large effect from the studies that used different tests for pre-test and post-test. So, using the same-test may have an effect on the results. However, heterogeneity in terms of I^2 does not differ as both groups have large heterogeneity. The p-value is less than .10, so the results are statistically significant.

Table 20

Subgroup Analysis of studies by the Subject Area

Subject area	k	*g*	*95% CI*	I^2	$p_{subgroup}$
Total	3	1.08	[-1.20, 3.37]	93%	.20
Mathematics	11	0.57	[-0.10, 1.23]	88%	
Science	4	1.10	[0.33, 1.86]	16%	
Language	2	0.31	[-3.02, 3.64]	15%	
Social Science	2	0.48	[-4.49, 5.44]	23%	

Table 20 shows that the number of studies that focused on mathematics achievement is larger than any other subject area. The largest effect size belongs to studies that focus on the scientific outcome. The studies focus on the total and mathematics academic achievements have the highest heterogeneity. The reason for mathematics achievement being heterogeneous can be the sample size. The reason for the heterogeneity of the total achievement may result from Booij's (2017) study as it was detected as an outlier and its large sample size. The p-value is larger than .10 which makes this statistically not significant.

Table 21below shows the summary results of the seven hypotheses testing at a glance. The detailed statistical results computed for each hypothesis are reported in Tables 16 – 22. Only hypothesis number 7 fails to reject. All hypothesis from 1 to 6 yielded statistically significant results.

Table 21

Summary Results of Hypothesis Testing

Null hypothesis	Reject	Fail to reject
1 There is no difference between the effect size of the studies that have a small sample size (n<100) and a large sample size (n>100).	X	
2 There is no difference between the study's effect sizes with different grade levels of students as their sample.	X	
3 There is no difference between the effect size of the studies that are written in English and Turkish.	X	
4 There is no difference between the effect size of the studies that are designed as one group pre-post test and control treatment groups.	X	
5 There is no difference between the effect size of the studies that are published in the journal and that are doctoral dissertations that are not published.	X	
6 There is no difference between the effect sizes of the studies that are using the same pre-test and post-test and different pre-tests and post-tests.	X	
7 There is no difference between the effect sizes of the studies that are investigating different subject areas		X

Testing Publication Bias

In meta-analysis, researchers ought to include all the research. Most often studies are missing from a meta-analysis. These are studies that were not published because they did not report favorable results. This causes publication bias. Firstly, it is advised to search grey literature to decrease publication bias. In this study we made sure to include the gray literature. As evidence this study included *half of the studies* were unpublished doctoral dissertations.

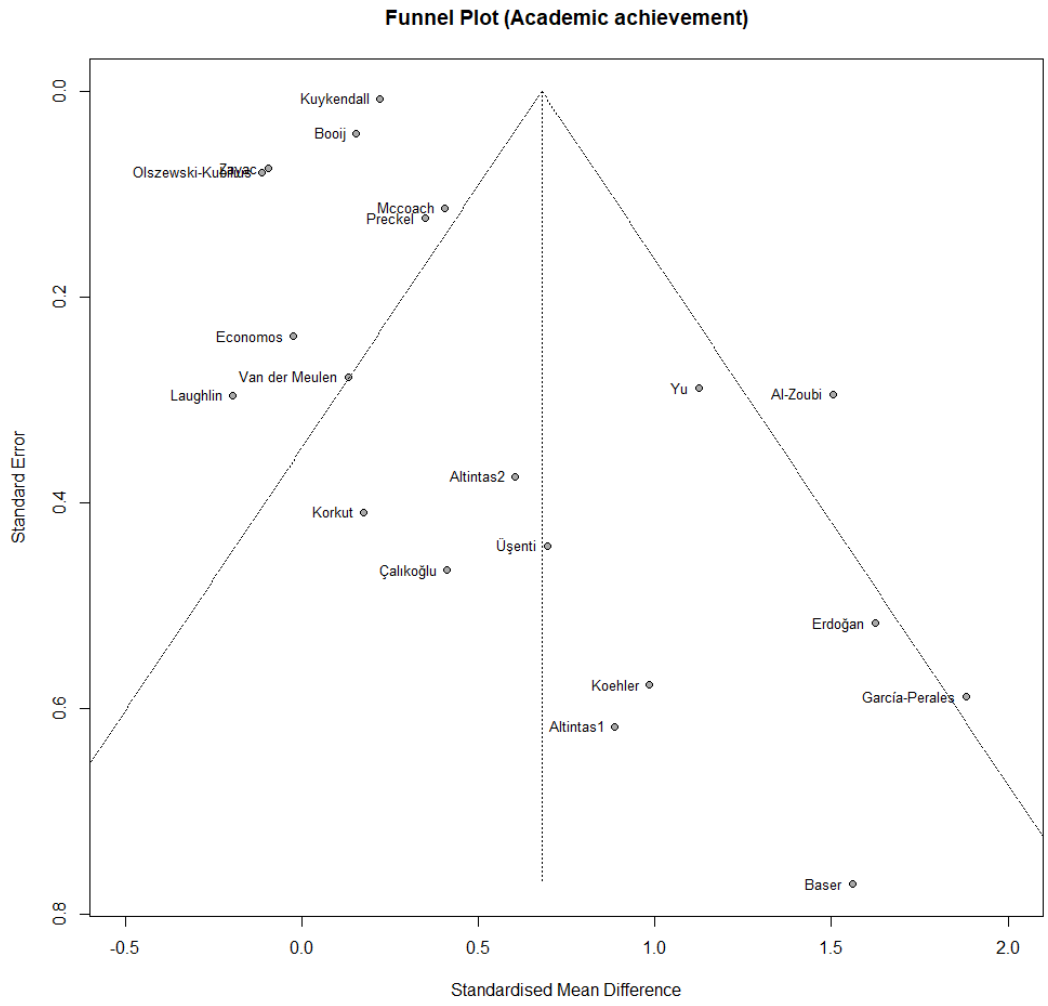
Then, there are statistical methods to check for the existence of publication bias. In this study, first funnel plot was drawn to observe possible publication bias then, Duval and Tweedie's Trim-and-Fill method was used to calculate the magnitude. This method is the suggested method because it can be interpreted by many researchers (Harrer et al., 2022, p.262).

Funnel Plot

Figure 9 is a pictorial representation showing publication bias by applying a statistical procedure that involves the sample size and standard error measures (Harrer et al., 2022, p.231).

Figure 9

Funnel Plot of all Included Studies



If there were no publication bias, then the output should be a symmetrical funnel. The distribution of the studies was not symmetric. The main concern is that the studies in the upper left corner, Booij (2017), Kuykendall (2020), Olszewski-Kubilius (2017), and Zayac (2013), are the most precise studies, but they do not follow the funnel pattern. They should be closer to the pooled effect size, but they were not. Then, the other studies, Economos (2021) and Laughlin (2013), did not following the pattern either, but this was not much of big concern as there is at least one article, Al-Zoubi (2014), that balances those two out. However, it is not possible to conclude publication bias directly by looking the asymmetry as it also occurs in other circumstances. For instance, if the studies have high between-study heterogeneity or low-quality studies show larger effect sizes the lead to asymmetry. Or high-quality studies which are more rigorous make the shape asymmetric as well (Harrer et al., 2022, p.262). We previously observed that the data have substantial heterogeneity. However, there is a possibility in our data that there is publication bias.

Duval & Tweedie Trim and Fill Method

As we observed that there is an asymmetry in funnel plot, we need to find a way to estimate the pooled effect size better. Duval and Tweedie Trim and Fill Method marks outliers in the funnel plot and then removes (hence the term trimmed) them. Then the pooled effect size is recalculated. After that, to make the shape symmetric, considering the new pooled effect size, additional effect sizes are included to balance the trimmed studies. Lastly, corrected pooled effect size is recalculated using all the studies.

The trim and fill method was be used twice with and without outliers since the heterogeneity in our study was substantial (Harrer et al., 2022, p.243). The first

one includes all the studies and the second one includes the studies except the outliers.

The trim and fill procedure added seven new effect sizes. These studies were added to balance some of the possible outliers that we detected before which are Erdoğan (2014), Karataş (2013), and Özçelik (2017). Also, there are several studies added to balance some studies that have sample size is small and have relatively high effect size which are Al-Zoubi (2014), Baser (2016), García-Perales (2019), and Yu (2020). After this trim and fill procedure shown in Table 22, the pooled effect size was calculated again as g=0.25 which is considered as small effect size. This outcome was expected as we assumed that studies with small samples with small effect sizes were missing and filled accordingly. All the new studies added to the calculation had negative effect sizes. For instance, to balance Yu which had g=1.13, the filled study included had effect size g=-0.70. However, this was not statistically significant finding as the p value was 0.28. One can observe the new funnel plot with filled studies in Figure 10 below.

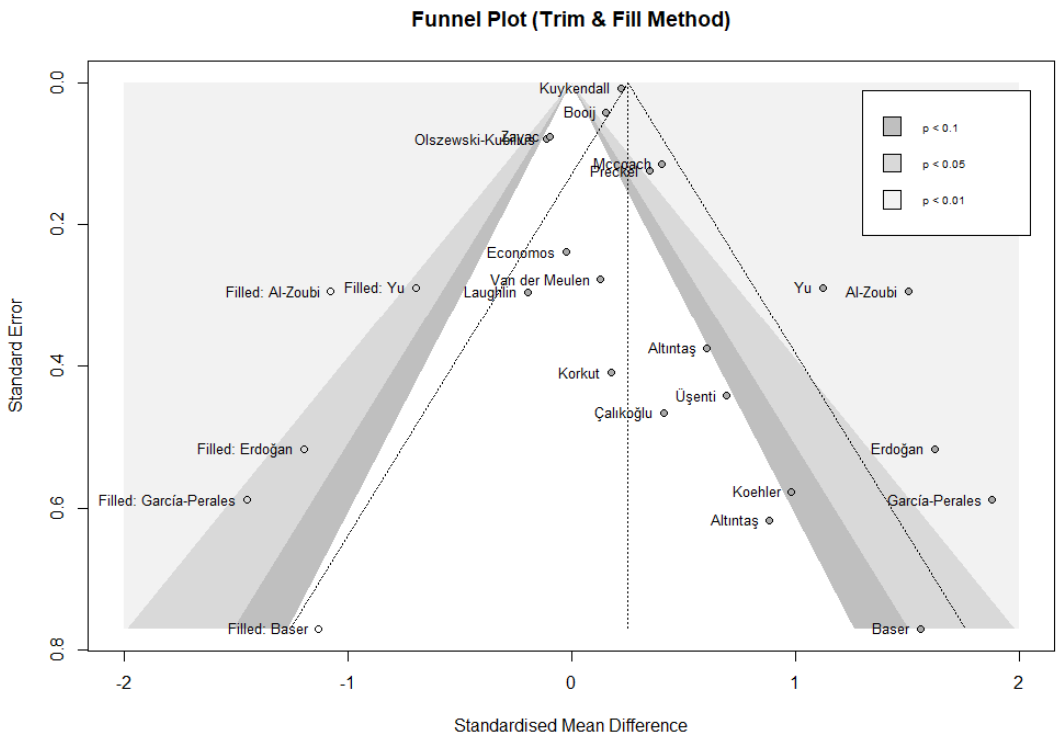
Table 22

Pooled Effect Size after Trim and Fill Procedure

SMD (g)	%95 CI	*p*
0.25	[-0.21, 0.72]	0.28

Figure 10

Trim and Fill Funnel Plot for all Studies



In this new figure, Figure 10, the studies Al-Zoubi (2014), Baser (2016), Erdoğan (2014), Garcia-Perales (2019), and Yu (2020) were outliers. All of these studies had large effect sizes. Therefore, to balance those, symmetric studies with respect to the new pooled effect size were added. However, again Figure 10 does not seem symmetric.

In the next step, the trim and fill procedure was applied after removing the studies that were considered as possible influential cases and outliers which are; Booij (2017), Olszewski-Kubilius (2017), Karataş (2013), Kuykendall (2020), Özçelik (2017), and Zayac (2013). This time six other studies were added to balance the studies: Al-Zoubi (2014), Baser (2016), Erdoğan (2014), García-Perales (2019), Koehler (2013), and Yu (2020). The results were shown in Table 23.

Table 23

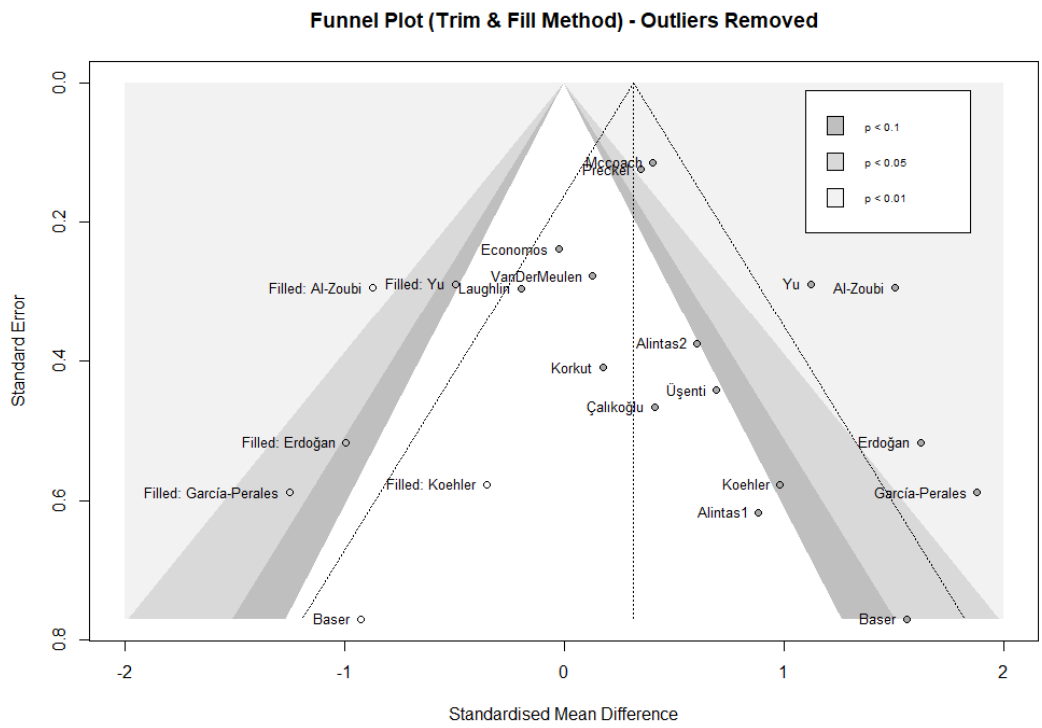
Pooled Effect Size after Trim and Fill Procedure without Outliers

SMD (g)	%95 CI	*p*
0.32	[-0.07, 0.70]	.10

It can be seen from the Table 23 that g=0.32 after the trim and fill procedure. It decreased again but not as much as the previous time. The new funnel plot after the removal can be seen in Figure 11 below. With the new filled effect sizes, the shape becomes more symmetric. Yet, it still does not look like a funnel.

Figure 11

Trim and Fill Funnel Plot for Studies except the Possible Outliers



CHAPTER 5: DISCUSSION

Introduction

This study focused on the academic achievement of gifted and talented students after receiving an intervention. The intervention that was of particular interest was enrichment. In order to compute the overall effect size and provide evidence regarding its robustness of future possible enrichment programs on academic achievement of talented and gifted students, a meta-analysis was employed. Throughout this study the overall effect size was reported in terms of Hedges' g along with its 95% confidence interval (see Chapter 3). To check the robustness of the findings, typical meta-analysis heterogeneity procedures were followed (see Figure 8). Finally, a publication bias determination test was conducted and add as an additional evidence for the robustness of the computed overall effect size.

In this chapter, first a brief overview of the study was presented, then the reported findings were discussed within the scope of the research, and lastly, recommendations for future studies were presented.

Overview of the Study

This research aimed to examine the relationship between enrichment programs and gifted and talented students' academic achievement after receiving an enrichment program. The only research question was, "What is the effect of the enrichment programs for gifted and talented students on their academic achievement?" The research method used in this study was meta-analysis, so; the data used were secondary data. That is the data came out from already conducted and

published scientific research. Using these data, effect size calculations for each study were done. Then as per meta-analysis routine the effect sizes were converted into Hedges' g to minimize small sample bias impact. The statistical model used in this meta-analysis process was a random-effects model (RMS) as the literature on meta-analysis recommends this model for possible variables affecting between-study heterogeneity. Several procedures were used to estimate the heterogeneity and to check the robustness of the computed overall effect size. Seven different subgroup analyses were explored to determine the study's possible moderators. The null hypotheses of these subgroup analyses were:

Null hypothesis 1: There is no statistical difference between the effect size of the studies that have a small sample size ($n<100$) and a large sample size ($n>100$).

Null hypothesis 2: There is no statistical difference between the study's effect size with different grade levels of students as their sample.

Null hypothesis 3: There is no statistical difference between the effect sizes of the studies that are written in English and Turkish.

Null hypothesis 4: There is no statistical difference between the effect size of the studies that are designed as one group pre-post test and control treatment groups.

Null hypothesis 5: There is no statistical difference between the effect size of the studies that are published in the journal and that are doctoral dissertations that are not published.

Null hypothesis 6: There is no statistical difference between the effect sizes of the studies that are using the same pre-test and post-test and different pre-tests and post-tests.

Null hypothesis 7: There is no statistical difference between the effect sizes of the studies that are investigating different subject areas.

Lastly, publication bias which was a great concern in the meta-analyses was tested using the funnel plot model along with trim and fill procedures.

Discussion of Major Findings

Referring to the statistical findings reported in the chapter 4, it may be possible to make some assumptions about the effectiveness of the enrichment programs on the academic achievement of gifted and talented students. There were 22 studies involving a total of 99186 observations in this meta-analysis. The overall effect size (SMD g=0.68) was considered a medium effect (see Figure 5). The value showed that the overall effect size favors the treatment group. This means that our result supports the current literature reports that enrichment programs positively affect gifted and talented students' academic achievement. However, not all the results of the studies showed a positive effect in favor of the gifted and talented students. There were studies in which the control group showed better results. The computed prediction interval was between negative 0.82 and positive 2.18 (see Table 9) suggesting that some future research may provide results favoring the control group (Harrer et al., 2022, p.152). After computing the overall effect size, statistical heterogeneity analyses were performed to check the robustness of the findings. For example, Cochran's Q test showed some statistically significant degree of heterogeneity between the studies as (Q=136.44, df =21; see Table 10). Therefore, we conducted further heterogeneity test to determine a better understanding of heterogeneity. I^2 values are reported widely in meta-analysis. In this study, the calculated I^2=83.9% refers to substantial heterogeneity (Harrer et al., 2022, p.146). A value that is neither dependent on the number of studies on the meta-analysis nor the precision of the studies is τ^2. Although its interval [0.24, 1.31] was all positive, it did not contain zero. This was another statically proof indicating between-study

heterogeneity existence (Harrer et al., 2022, p.151). These findings on heterogeneity created the need to explore further.

One cause of heterogeneity might be due to a study or several studies in the data set might have huge effect on the overall effect size. Such studies are referred to outliers. The influential cases, on the other hand, may align with the overall effect size but have a large impact on it. The results of these possible outliers deviate from the overall effect size significantly so that they can influence or misrepresent the true effect size. So, we investigated via the forest plot (see Figure 5) and detected potential studies that may be considered outliers and/or influential. These studies were

1. Booij (2017) [5],
2. Kuykendall (2020) [13],
3. Mccoach (2014) [15],
4. Olszewski-Kubilius (2017) [16],
5. Preckel (2019) [18], and
6. Zayac (2013) [22].

The second observation was two studies were considered far from this base line. These were Karataş (2013) (SMD g =2.72) and Özçelik (2017) (SMD g =2.75). These two effect size estimates were four times larger than our computed overall effect size. An unlikely result to be obtained through an enrichment program.

Three statistical methods were used to detect and/or justify these possible outliers and influential cases. These were brute force, leave-one-out, and graphic display of study heterogeneity (GOSH) respectively.

The first analysis used brute force method and it flagged Mccoach (2014) [15] and Preckel (2019) [18]. So, only two studies ([15], [18]) out of the eight studies

determined by the forest plot were accounted for. After removing these two studies from the dataset, the overall effect size was recalculated, and the results did not statistically change. The effect size dropped to g=0.64 from g=0.68, but the confidence interval shrank as well, therefore Preckel (2019) and Mccoach (2014) could not be classified as outliers. The recomputed prediction interval again justified enrichment programs in favor of the gifted and talented students for future studies (see Table 11). The brute force method did not account any explanations regarding the other six studies mentioned above. So, two other major methods were conducted: the leave-one-out and GOSH method.

The leave-one-out method used the Baujat plot and it flagged Al-Zoubi (2014) [1], Karataş (2013) [10], Olszewski-Kubilius (2017) [16], Özçelik (2017) [17], and Zayac (2013) [22]; however, these studies almost have the same influence on the overall effect size as the rest of the studies. On the other hand, Kuykendall (2020) does not contribute much to the heterogeneity but has a large impact on the overall result.

The GOSH diagnostics (Olkin et al., 2012) is an intensive computational method that iterates through all possible subsets, in our case 4,194,303 unique sub-cases. Our GOSH plot analysis shows a unimodal, compact graph with high heterogeneity. The graph did not show any signs of clusters (i.e., unique subsets) in the data. So, the Figure 7 shows that if we take different subsets of our data, the effect size value ranges between 0.25 and 1.0. Also, the heterogeneity can take any value. However, there is an accumulation between 50% and 100%.

So far, all our findings support the assumptions about the effectiveness of the enrichment programs in favor on the academic achievement of gifted and talented students.

In addition to the overall findings, to see whether any subgroups existed in the data, subgroup analyses were conducted. The pre-determined variables were:

1. The studies' sample size,
2. Grade level,
3. English vs. Turkish,
4. Study design,
5. Publication type,
6. Same pre-post vs. parallel or different post-test, and
7. Subject areas.

The first analysis showed that the sample size is a possible subgroup for this study. The overall effect size of the studies with less than 100 is significantly larger (SMD g=0.97), whereas the effect size of large sampled studies was (SMD g=0.14). This result made perfect sense since teaching to a smaller number of students with a specific purpose would yield higher mean difference scores when compared to an intervention with much larger number of students, like those statewide initiatives. Also, more resources (time, money) are allocated for the studies that involve larger samples. Therefore, the researchers have the opportunity to conduct more detailed work. For instance, Baser (2016) mentioned that it was difficult to reach gifted and talented students as there were no special schools in the area, and the identification of the students needed resources. So, the larger sampled studies may result closer estimates to the overall effect size which may create heterogeneity. Additionally, the criteria for giftedness may be looser in larger sampled studies since it is generally hard to reach a high number of gifted and talented participants. Therefore, one needs to further check whether there is such a connection in the literature.

For the next analyses, three grade groups were compared. K-5 (g=0.79) and 6-8 (g=0.72). Our results showed that younger pupils benefit more from enrichment programs when compared to high-school students. There were also only three studies conducted with grades 9-12, which is very small. So, we clearly need more numbers of studies to validly establish the magnitude and direction of enrichment programs on high school students. Perhaps more studies could change the SMD g value to comparable magnitudes as the other two levels.

The next analysis investigates language as a subgroup. Studies published and conducted in English, SMD g =0.42, whereas, in Turkish, it was SMD g=1.28. This outcome is interesting. All studies conducted in Turkish had a small sample size. There were only seven studies, none of which has large sample size. So, we clearly need more numbers of studies with a higher sample size to validly establish the magnitude and direction of Turkish enrichment programs.

The following analysis focused on the design of the studies. There is again a difference as the studies with an experimental-control design have a medium effect size (g=0. 61), whereas one group pre-test and post-test have a large effect (g=1.00). However, again the number of the studies that used one group pre-test and post-test studies is only four. More studies are needed to make clearer inferences.

After that, the publication type divided into four categories was checked. It can be seen that there is only one study that is neither a journal article nor a thesis. There is also only one master's thesis. So, the two categories we focus on journal articles and doctoral dissertations. Although they have similar confidence intervals, the overall effect size of the journal articles (g=0.64) was considered a medium effect, whereas doctoral dissertations have a large effect (g=0.85). This was not expected as the published articles generally have a larger effect size due to

publication bias (McAuley et al., 2000, as cited in Harrer et al., 2022, p.230). We can think that doctoral dissertations may have publication bias as well due to the pressure to publish their research.

The last possible subgroup was whether the studies used the same tests for pre-test and post-test. These results were interesting as there were almost the same amount of studies in each group. Six out of seven Turkish studies used the same test as pre and posttests. So, the results somehow aligned with the language analysis. The studies that used the same test had a high effect size (g=1.23), whereas the studies that used different tests had a small effect size (g=0.17). The studies that use the same parallel tests might face reliability problems. Another factor might be retention. The students might remember the questions and answers of the pre-tests, which may be why the effect size is that much larger.

The last analysis explores different subject areas. It was found that most studies focused on mathematics achievement. The effect size of mathematics achievement (g=0.57) was close to overall effect size. In the social science area, the effect size (g=0.48) was similar but the sample size is very small, only two. Studies focused on total, language, and science achievements also had small sample sizes. Among them, four studies focused on science had a large effect size (g=1.10). In order to make better inferences more research had to be done in academic areas other than mathematics.

Lastly, publication bias was computed using a funnel plot. It was observed that the plot was not symmetric. Therefore, one can conclude that there was possibility of publication bias in the sample of this research.

Overall Conclusion

There were other meta-analyses on the same topic. The earliest one was conducted by Wallace (1989). In this meta-analysis, the studies included were published between the years of 1931 and 1986 (over a period of 55 years). There were twenty studies in the meta-analysis. My meta-analysis (2013-2021) sample size was 22. Besides showing the fast expansion of the literature on enrichment programs regarding gifted and talented students, it is also a sign that gifted education practices are still a concern for many researchers. Wallace (1989) published a positive overall effect size of 0.57. However, he failed to report any heterogeneity measure, so it is not possible to compare the heterogeneity findings.

Another, more recent meta-analysis on the same issue was conducted by Kim (2016). The studies included in her meta-analysis were published between the years of 1985 and 2014. Her take on academic achievement definition was broader as she also included outcomes such as thinking processes or cognitive skills. She published a large positive effect (g=0.96) under the random-effects model. Also, Cochran's Q value was reported in her study as Q (13) =85.32. This result is as well a high heterogeneity. She failed to report on any subgroup analysis and publication bias. Unlike my study, she used commercial software that made it impossible for us to validate her results.

Our results on overall effect size and heterogeneity seem to be of the same direction as the two previous studies mentioned above. In terms of magnitude it is closer to Wallace’s. One unique attempt of this study was to report and provide further evidence regarding heterogeneity. The Table 24 below is a summary depicting the possible outliers and influential cases. These were studies numbered 1,5,10,13,16,17, and 22.

Table 24

Investigating Possible Outliers

Study number	H1	H2	H3	H4	H5	H6	H7	*g*
1	small	6-8	Eng	pre-post	Journal	yes	total	1,51
5	large	9-12	Eng	exp-cont	Non-Journal	no	total	0,15
10	small	K-5	Tur	exp-cont	Dissertation	yes	math	2,72
13	large	K-5	Eng	exp-cont	Dissertation	no	math	0,22
16	large	K-5	Eng	exp-cont	Journal	no	math	-0,11
17	small	K-5	Tur	exp-cont	Dissertation	yes	math	2,75
22	large	6-8	Eng	exp-cont	Dissertation	no	math	-0,1

By looking at the Table 24, we observed that studies 5, 13, 16, and 22 have a larger sample size than the other studies, which made them influential cases. The other studies numbered 1, 10, and 17 have a relatively smaller sample size, but they still deviated from the pooled effect size, which made them influential cases as well. Interestingly, all three of those studies used the exact same pre-test and post-test to measure the students' academic outcomes. Hence, the large SMD g size value can be attributed to possible student retention when answering the same test as a posttest.

Finally, we have made the dataset and codes available for anyone interested in our research for reliability and validity purposes.

Recommendations for Future Research

Including our study, three other meta-analyses on the academic achievement of enrichment programs offered to gifted and talented students were reported. Our results were in line with the two previous ones. Furthermore, we did a subgroup analysis indicating that the results of large sample size studies contribute to the precision of the meta-analysis. So, we recommend studies be conducted with larger student samples.

Student test retention seems to play an important role resulting in high SMD g values. Since we deal with gifted and talented students, we recommend utilizing a parallel test form for the post-test assessment. If there is no such possibility, then we recommend including a third measurement set that considers test retention.

The result of our meta-analysis showed that language is a factor to be taken into account. The number of Turkish studies is low; more studies need to be conducted so that the reported SMD g value becomes robust. As it stands, the Turkish SMD g-value seems too high and is not very plausible. Also, almost all the Turkish studies used the same test as pre-test and post-test. Additionally, all of the studies published in Turkish had a sample size of less than 100. In order to see whether these variables affect the overall effect size, more studies need to be published in Turkish with a larger sample size and parallel tests.

As our literature review showed, there is an increasing trend in enrichment programs that pertain to gifted and talented students. In order to increase the sample pool of a future meta-analysis, we urge academic journals to require authors to include explicit data such as mean, standard deviation, and sample size or effect size. As a result, this would increase the pool for meta-analysis.

Our meta-analysis was a mean difference type. Other types of meta-analysis, such as correlational or combined, should also be conducted. This would help to increase the degree of validation of the generalization of the effect size. Moreover, a more comprehensive method called meta-regression could be used to explain the data better. By doing so, more than one variable in a subgroup can be utilized, and their interaction can be explained (Harrer et al., 2022, p.198).

REFERENCES

Al-Zoubi, S. M. (2014). Effects of enrichment programs on the academic achievement of gifted and talented students. *Journal for the Education of Gifted Young Scientists, 2*(2), 22-27.

Altıntaş, E. (2014). *Üstün zekâlı öğrenciler için yeni bir farklılaştırma yaklaşımının geliştirilmesi ve matematik öğretiminde uygulanması* (Publication No. 372284) [Doctoral dissertation, Marmara University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Altıntaş, E., & Özdemir, A. Ş. (2015). The effect of the developed differentiation approach on the achievements of the students. *Eurasian Journal of Educational Research*, (61).

Astin, A., & Ross, S. (1960). Glutamic acid and human intelligence. *Psychological Bulletin, 57*, 429-434.

Atılgan, H. (2018). Türkiye'de kademeler arası geçiş: Dünü-bugünü ve bir model önerisi. *Ege Eğitim Dergisi*, *19*(1), 1-18.

Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education, 5*(6)

Baujat, B., Mahé, C., Pignon, J. P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine, 21*(18), 2641-2652.

Baum, S. M., Renzulli, J. S., & Hébert, T. P. (1995). Reversing underachievement: Creative productivity as a systematic intervention. *Gifted Child Quarterly*, *39*(4), 224-235.

Baser, C. (2016). *Teaching gifted and talented children by applying enrichment and acceleration programs (Teaching math and science in primary schools in Iraq)* [Unpublished doctoral dissertation]. International Black Sea University.

Battiste, M. (2011). Maintaining Aboriginal identity, language, and culture in modern society. In M. Battiste (Ed.), *Reclaiming indigenous voice and vision* (pp. 192-208). Vancouver: University of British Columbia Press.

Booij, A. S., Haan, F., & Plug, E. (2017). Can gifted and talented education raise the academic achievement of all high-achieving students? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2988183

Boli, J., Ramirez, F. O., & Meyer, J. W. (1985). Explaining the origins and expansion of mass education. *Comparative Education Review, 29*(2), 145-170.

Brandon, L., Reis, S., & McGuire, C. (2021). Perceptions of talented university students related to opportunities and autonomy for creative productivity. *Gifted Education International*, *37*(3), 217-240.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 16-33). Cambridge University Press.

Campbell, J. R., Wagner, H., & Walberg, H. J. (2000). Academic competitions and programs designed to challenge the exceptionally talented. In K. A. Heller (Ed.), *International handbook of giftedness and talent* (pp.523-535). Elsevier Science & Technology.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87*(3), 442-449.

Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, *96*(3), 443–456.

Cuijpers, P. (2016). *Meta-analyses in mental health research. A practical guide.* Amsterdam: Vrije Universiteit Amsterdam.

Çalıkoğlu, B. S. (2014). *Üstün zekâlı ve yetenekli öğrencilerde derinlik ve karmaşıklığa göre farklılaştırılmış fen öğretiminin başarı, bilimsel süreç becerileri ve tutuma etkisi* (Publication No. 356628) [Doctoral dissertation, İstanbul University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Dai, D. Y., & Chen, F. (2021). *Paradigms of gifted education: A guide to theory–based, practice–focused research*. Routledge.

Dai, D. Y., & Coleman, L. J. (2005). Epilogue: Conclusions and implications for gifted education. *Journal for the Education of the Gifted 28*(3–4), 374-388.

Davis, G. A., Rimm, S.B., & Siegle, D.B. (2010). *Education of the gifted and talented.* (6th ed.). Pearson.

Desjardins, R. (2015). Education and social transformation. *European Journal of Education*, *50*(3), 239–244. https://doi.org/10.2307/26609272

Economos, N. (2021). *Mathematics achievement among high-ability second grade students* (Publication No. 28716949) [Doctoral dissertation, Wilmington University]. ProQuest Dissertations & Theses Global.

Erdoğan, S. C. (2014). *Bilimsel yaratıcılığı temel alan farklılaştırılmış fen ve teknoloji öğretiminin üstün zekalı ve yetenekli öğrencilerin başarı, tutum ve yaratıcılığına etkisi* (Publication No. 356630) [Doctoral dissertation, İstanbul University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (Pico) as a search strategy tool on literature search quality: A systematic review. *Journal of the Medical Library Association, 106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Esterhuizen, T. M., & Thabane, L. (2016). Con: Meta-analysis: Some key limitations and potential solutions. *Nephrology Dialysis Transplantation, 31*(6), (882–885). Oxford University Press. https://doi.org/10.1093/ndt/gfw092

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist 33*(5). https://doi.org/10.1037/0003-066X.33.5.517.a

Fagard, R. H., Staessen J. A., & Thijs, L. (1996). Advantages and disadvantages of the meta-analysis approach. *Journal of Hypertension, Supplement, 14*(2). https://doi.org/10.1097/00004872-199609002-00004

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st Century. *Journal of Clinical Epidemiology, 48*(1), 71-79.

Feldhusen, J. F. (1980). *The three-stage model of course design*. Educational Technology Publications.

Feldhusen, J. F., & Kolloff, M. B. (1988). A three-stage model for gifted education. *Gifted Child Today Magazine, 11*(1), 14-20.

Feldhusen, J. F., & Ruckman, D. R. (1988). A guide to the development of Saturday programs for gifted and talented youths. *Gifted Child Today Magazine, 11*(5), 56-61.

García-Perales, R., & Almeida, L. S. (2019). An enrichment program for students with high intellectual ability: Positive effects on school adaptation. *Comunicar*. *Media Education Research Journal, 27*(60), 39–47.

Getzels, J. W., & Jackson, P. W. (1958). The meaning of “giftedness”: An examination of an expanding concept. *The Phi Delta Kappan, 40*(2), 75-77.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Godwin, K. A., & Altbach, P. G. (2016). A historical and global perspective on liberal arts education: What was, what is, and what will be. *International Journal of Chinese Education, 5*(1), 5-22.

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *The Journal of Educational Research*, *83*(6), 313-327.

Gonzalez-Mulé, E., & Aguinis, H. (2018). Advancing theory by assessing boundary conditions with metaregression: A critical review and best-practice recommendations. *Journal of Management, 44*(6), 2246–2273.

Grant, B. A., & Piechowski, M. M. (1999). Theories and the good: Toward child-centered gifted education. *Gifted Child Quarterly*, *43*(1), 4-12.

Gubbins, E. J., McCoach, D. B., Foreman, J. L., Gilson, C. M., Bruce-Davis, M. N., Rubenstein, L. D., ... & Waterman, C. (2013). *What works in gifted education mathematics study: Impact of pre-differentiated and enriched curricula on general education teachers and their students.* National Research Center on the Gifted and Talented.

Gurevitch J, Koricheva J, Nakagawa S, Stewart G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*(7695), 175–182.

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval

qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Hansen, C., Steinmetz, H., & Block, J. (2021). How to conduct a meta-analysis in eight steps: a practical guide. *Management Review Quarterly, 72*(1), 1-19 Springer Science and Business Media Deutschland GmbH.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2022). *Doing meta-analysis with R: A hands-on guide.* CRC Press.

Harris, D. J., Allen, K. L., Vine, S. J., & Wilson, M. R. (2021). A systematic review and meta-analysis of the relationship between flow states and performance. *International Review of Sport and Exercise Psychology*, 1-29.

Herrnstein, R. J., & Murray, C. (1996). *The bell curve: Intelligence and class structure in American life.* (Illustrated ed.). Free Press.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

İş Güzel, Ç. (2013). Eğitim sistemimizdeki ölçme ve değerlendirme uygulamaları nasıl olmalıdır?. *Cito Eğitim: Kuram ve Uygulama*, (*21)*, 10-16.

Karataş, Y. (2013). *Farklılaştırılmış matematik öğretiminin üstün zekâlı ve yetenekli öğrencilerde erişiye, yaratıcılığa, tutuma ve akademik benliğe etkisi* (Published No. 370186) [Doctoral dissertation, İstanbul University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Kent, S. D. (1992). *The effects of acceleration on the social and emotional development of gifted elementary students: A meta-analysis* (Publication No.

9316362) [Doctoral dissertation, University of Georgia]. ProQuest Dissertation & Theses Global.

Kim, M. (2016). A meta-analysis of the effects of enrichment programs on gifted students. *Gifted Child Quarterly*, *60*(2), 102–116. https://doi.org/10.1177/0016986216630607

Kitano, M.K. (1998). Diverse gifted women. In J. H. Borland, (Ed.), *Rethinking gifted education*. Teachers College Press.

Klein, A. G. (2000). Fitting the school to the child: *The mission of Leta Stetter Hollingworth, founder of gifted education. Roeper Review*, *23*(2), 97-103.

Koçak, Ö., Göksu, İ., & Göktaş, Y. (2021). The factors affecting academic achievement: A systematic review of meta analyses. *International Online Journal of Education and Teaching*, *8*(1), 454-484.

Koehler, K. E. (2013). *Challenging gifted and talented high school students using a cognitive apprenticeship to build historical expertise* (Publication No. 3567763) [Doctoral dissertation, Northern Illinois University]. ProQuest Dissertations & Theses Global.

Korkut, Ş. (2017). *Üstün yetenekli öğrencilerin eğitiminde bütünleştirilmiş müfredat modeline göre farklılaştırılmış sosyal bilgiler öğretimi* (Publication No. 480346) [Master's thesis, İnönü University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Krahenbuhl, K. S. (2016). Student-centered education and constructivism: Challenges, concerns, and clarity for teachers. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *89*(3), 97–105.

Kuykendall, T. M. (2020). *Best match: Evaluating the impact of service models on the math achievement of culturally different gifted elementary learners*

(Publication No. 15971023893545) [Doctoral dissertation, Cleveland State University]. Ohio LINK Electronic Theses and Dissertations Center.

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19*(3), 415–428. https://doi.org/10.3102/00028312019003415

Laughlin, S. R. (2013). *Robotics: Assessing its role in improving mathematics skills for grades 4 to 5* (Publication No. 3557417) [Doctoral dissertation, Capella University]. ProQuest Dissertations & Theses Global.

Liberati, A. (1995). "Meta-analysis: statistical alchemy for the 21st century": Discussion. A plea for a more balanced view of meta-analysis and systematic overviews of the effect of health care interventions. *Journal of Clinical Epidemiology, 48*(1), 81-86.

Lindholm-Leary, K., & Borsato, G. (2006). Academic achievement. In *Educating English Language Learners* (pp. 176-223). Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Sage Publications, Inc.

McAuley, L., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?. *The Lancet, 356*(9237), 1228-1231.

McCoach, D. B., Gubbins, E. J., Foreman, J., Rubenstein, L. D. V., & Rambo-Hernandez, K. E. (2014). Evaluating the efficacy of using predifferentiated and enriched mathematics curricula for Grade 3 students: A multisite cluster-randomized trial. *Gifted Child Quarterly, 58*(4), 272–286. https://doi.org/10.1177/0016986214547631

MEB. (2017). *Bakan Yılmaz, yeni müfredatı açıkladı*. T.C. Millî Eğitim Bakanlığı. Retrieved from https://meb.gov.tr/bakan-yilmaz-yeni-mufredati-acikladi/haber/14112/tr

MEB. (2020). *Bilim ve sanat merkezleri güçleniyor*. T.C. Millî Eğitim Bakanlığı. Retrieved from https://www.meb.gov.tr/bilim-ve-sanat-merkezleri-gucleniyor/haber/21827/tr

Meyer, A. E. (1965). *An educational history of the western world*. New York: McGraw-Hill.

Milli Eğitim Temel Kanunu, Resmi Gazete, 14574, (1973).

Moon, S. M., Feldhusen, J. F., Powley, S., Nidiffer, L., & Whitman, M. W. (1993). Secondary applications of the Purdue three stage model. *Gifted Child Today Magazine*, *16*(3). https://doi.org/10.1177/107621759301600301

Neuenschwander, M. P., & Garrett, J. L. (2008). Causes and consequences of unexpected educational transitions in Switzerland. *Journal of Social Issues 64*(1), 41-58.

Oğurlu, U. (2021). Overview of meta-analyses on giftedness. *Gifted and Talented International*, *35*(2), 110–127. https://doi.org/10.1080/15332276.2021.1893135

Olkin, I., Dahabreh, I. J., &Trikalinos, T. A. (2012). GOSH - a graphical display of study heterogeneity. *Research Synthesis Methods*, *3*(3), 214–223.

Olszewski-Kubilius, P., Steenbergen-Hu, S., Thomson, D., & Rosen, R. (2017). Minority achievement gaps in STEM: Findings of a longitudinal study of project Excite. *Gifted Child Quarterly*, *61*(1), 20–39. https://doi.org/10.1177/0016986216673449

Özçelik, T. (2018). *Üstün yetenekli öğrencilere yönelik geliştirilen farklılaştırılmış matematik dersi öğretim programının etkililiği* (Publication No. 484072) [Doctoral dissertation, Hacettepe University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Petitti, D. B. (2001). Approaches to heterogeneity in meta-analysis. *Statistics in medicine*, *20*(23), 3625-3633.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Pfouts, D. K., & Schultz, R. A. (2003). The benefits of for outdoor learning centers for young gifted learners. *Gifted Child Today*, *26*(1). https://doi.org/10.4219/gct-2003-91

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, *90*(1), 24-46.

Plucker, J. A., & Callahan, C. M. (2020). The evidence base for advanced learning programs. *Phi Delta Kappan*, *102*(4), 14-21.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A Product from the ESRC Methods Programme Version*, *1*(1), 92.

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., Scherrer, V., & Schneider, W. (2019). High-ability grouping: Benefits for gifted students' achievement development without costs in academic self-Concept. *Child Development*, *90*(4), 1185–1201.https://doi.org/10.1111/cdev.12996

Pritchard, A., & Woollard, J. (2010). *Psychology for the Classroom: Constructivism and Social Learning* (1st ed.). Routledge. https://doi.org/10.4324/9780203855171

Renzulli, J. S. (2021). The enrichment triad model: A guide for developing defensible programs for the gifted and talented. In *Reflections on Gifted Education: Critical Works by Joseph S. Renzulli and Colleagues* (pp. 193–210). Routledge.

Renzulli, J. S., & Reis, S. M. (2010). The schoolwide enrichment model: A focus on student strengths and interests. *Gifted Education International*, *26*(2–3). https://doi.org/10.1177/026142941002600303

Roberts, J. L., Inman, T. F., & Robins, J. H. (Eds.). (2022). *Introduction to gifted education*. Taylor & Francis.

Roeper, A. (1990). *Educating children for life.*Trillium Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* SAGE Publications, Inc.

Sak, U., Ayaş, B., Sezerel, B. B., Öpengin, E., Özdemir, N. N., & Demirel-Gürbüz, Ş. (2015). Türkiye'de üstün yeteneklilerin eğitiminin eleştirel bir değerlendirmesi.*Türk Üstün Zekâ ve Eğitim Dergisi, 5*(2), 110-132

Sapon-Shevin M. (1996) Beyond gifted education: Building a shared agenda for school reform. In J. H. Borland, (Ed.), *Rethinking gifted education*. Teachers College Press.

Schäfers, M. S. & Wegner, C. (2020). Long-term effects of the enrichment program "Kolumbus Kids". *Journal of Gifted Education and Creativity, 7*(2), 73-82.

Snyder, K. E., Fong, C. J., Painter, J. K., Pittard, C. M., Barr, S. M., & Patall, E. A. (2019). Interventions for academically underachieving students: A systematic review and meta-analysis. *Educational Research Review*, *28*, 100294.

Sönmez Çalış, Ö. A. (2019). *Türkiyeye tersine beyin göçü: Nitel bir araştırma* (Publication No. 583302) [Doctoral dissertation, Sakarya University]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Steenbergen-Hu, S., & Moon, S. M. (2011). The effects of acceleration on high-ability learners: A meta-analysis. *Gifted Child Quarterly*, *55*(1), 39–53.

Stromquist, N. P. (2002). *Education in a globalized world: The connectivity of economic power, technology, and knowledge*. Rowman & Littlefield.

Taylor, G.R. (2005). *Integrating quantitative and qualitative methods in research*. Oxford: University Press of America.

Treffinger, D. J., & Selby, E. C. (2009). Levels of service: A contemporary approach to programming for talent development. *Systems and Models for Developing Programs for the Gifted and Talented*, 629-654.

Treffinger, D.J., Nassab, C.A., & Selby, E.C. (2008). *Programming for talent development: Expanding horizons for gifted education.* The Routledge International Companion to Gifted Education.

Üşenti, Ü. A. (2013). Farklılaştırılmış Türkçe öğretim uygulamalarının üstün zekâlı ve yetenekli öğrencilerdeki bilişsel beceri ve başarı durumlarına etkisi. *International Journal of Social Science Research*, *2*(2).

Van der Meulen, R. T., Van der Bruggen, C. O., Spilt, J. L., Verouden, J., Berkhout, M., & Bögels, S. M. (2014). The pullout program day a week school for gifted children: Effects on social–emotional and academic functioning. *Child & Youth Care Forum*, *43*(3), 287-314. Springer US.

Vaughn, V. L., Feldhusen, J. F., & Asher, J. W. (1991). Meta-Analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly, 35*(2), 92–98

Wallace, T. A. (1989). *The effects of enrichment on gifted students: A quantitative synthesis* (Publication No. 9003748) [Doctoral dissertation, University of Illinois at Chicago]. ProQuest Dissertations & Theses Global.

Warne, R. T. (2016). *Publication bias currently makes an accurate estimate of the benefits of enrichment programs difficult: A postmortem of two meta-Analyses using statistical power analysis* (ED572464). ERIC. https://files.eric.ed.gov/fulltext/ED572464.pdf

Watzlawik, M., & Born, A. (Eds.). (2007). *Capturing identity: Quantitative and qualitative methods*. University Press of America.

West, R. R. (1993). A look at the statistical overview (or meta-analysis). *Journal of the Royal College of Physicians of London, 27*(2), 111.

Wissehr, C., Concannon, J., & Barrow, L. H. (2011). Looking back at the Sputnik era and its impact on science education. *School Science and Mathematics, 111*(7).

Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixson, D. D. (2019). Gifted students. *Annual Review of Psychology, 70*, 551-576.

Yu, H. P., & Jen, E. (2020). Integrating nanotechnology in the science curriculum for elementary high-ability students in Taiwan: Evidenced-based lessons. *Roeper Review, 42*(1), 38–48. https://doi.org/10.1080/02783193.2019.1690078

Zayac, J. M. (2013). The impact of elementary gifted mathematics programming: Moving into middle school research using the ecls-k database (Publication

No. 3604451) [Doctoral dissertation, Notre Dame of Maryland]. ProQuest Dissertations & Theses Global.

Appendix A

Data Set of the Included Article

Author	Pub. Date	Sample Size	Grade Level	Language	Design	Pub. Type
Al-Zoubi	2014	60	2	English	pre-post	Journal
Altıntaş	2015	30	2	English	exp-cont	Journal
Altıntaş	2014	12	1	Turkish	exp-cont	Dissertation
Baser	2016	10	1	English	pre-post	Dissertation
Booij	2017	2423	3	English	exp-cont	Non-Journal
Çalıkoğlu	2014	19	1	Turkish	exp-cont	Dissertation
Economos	2021	71	3	English	exp-cont	Dissertation
Erdoğan	2014	21	2	Turkish	exp-cont	Dissertation
García-Perales	2019	18	1	English	exp-cont	Journal
Karataş	2013	24	1	Turkish	exp-cont	Dissertation
Koehler	2013	14	3	English	exp-cont	Dissertation
Korkut	2017	24	1	Turkish	exp-cont	Masters
Kuykendall	2020	93891	1	English	exp-cont	Dissertation
Laughlin	2013	46	1	English	exp-cont	Dissertation
Mccoach	2014	332	1	English	exp-cont	Journal
Olszewski-Kubilius	2017	948	1	English	exp-cont	Journal
Özçelik	2017	28	1	Turkish	exp-cont	Dissertation
Preckel	2019	265	2	English	exp-cont	Journal
Üşenti	2013	22	1	Turkish	exp-cont	Dissertation
Van der Meulen	2013	52	1	English	pre-post	Journal
Yu	2020	56	1	English	pre-post	Journal
Zayac	2013	820	2	English	exp-cont	Dissertation

Appendix B

R Codes Used for Analysis

1. Pooling Effect Size

```
library(meta)
library(dmetar)
library(meta)
data(academic_achievement)
m.gen<- metacont(n.e = n.e,
    mean.e = mean.e,
    sd.e = sd.e,
    n.c = n.c,
    mean.c = mean.c,
    sd.c = sd.c,
    studlab = author,
    data = academic_achievement,
    sm = "SMD",
    method.smd = "Hedges",
    fixed = FALSE,
    random = TRUE,
    method.tau = "REML",
    hakn = TRUE,
    title = "Academic Achievement")
m.gen<- update.meta(m.gen, prediction = TRUE)
summary(m.gen)
```

2. Determining Possible Outliers Using Brute Force

```
find.outliers(m.gen)
```

3. Influence Analysis

```
m.gen.inf <- InfluenceAnalysis(m.gen, random = TRUE)
```

4. Plotting Influence Analysis

```
plot(m.gen.inf, "baujat")
plot(m.gen.inf, "influence")
plot(m.gen.inf, "es")
plot(m.gen.inf, "i2")
```

5. Gosh Plot

```
library(metafor)
m.rma<- rma(yi = m.gen$TE,
            sei = m.gen$seTE,
            method = m.gen$method.tau,
            test = "knha")
res.gosh<- gosh(m.rma)
```

Appendix B (Continued)

```
plot(res.gosh, alpha = 0.01)
res.gosh.diag<- gosh.diagnostics(res.gosh,
                                 km.params = list(centers = 2),
                                 db.params = list(eps = 0.08,
                                                  MinPts = 50))
res.gosh.diag
plot(res.gosh.diag)
library(magrittr)
update.meta(m.gen, exclude = c(3, 4, 16)) %>%
  summary()
```

6. Forest Plot

```
forest.meta(m.gen,
            sortvar = TE,
            print.tau2 = FALSE,
            leftlabs = c("Author", "g", "SE"))
forest.meta(m.gen,
            sortvar = TE,
            print.tau2 = FALSE,
            leftcols = c("studlab", "TE", "seTE", "CODED VARIABLE"),
            leftlabs = c("Author", "g", "SE", "sample"))
```

7. Subgroup Analysis

```
update.meta(m.gen,
        subgroup = SUBGROUPS,
        tau.common = FALSE)

update.meta(m.gen, subgroup = RiskOfBias, tau.common = TRUE)
```

8. Funnel Plot

```
library(meta)
funnel.meta(m.gen,
            xlim = c(-0.5, 2),
            studlab = TRUE)
        title("Funnel Plot (Academic Achievement)")
```

9. Trim and Fill

```
tf <- trimfill(m.gen)
```

Appendix B (Continued)

```
        tf.no.out <- trimfill(update(m.gen,
                        subset = -c(number of outlier1, number of
outlier2)))
        summary(tf)
        summary(tf.no.out)
```

10. contour <- c(0.9, 0.95, 0.99)

```
        col.contour <- c("gray75", "gray85", "gray95")
        ld <- c("p < 0.1", "p < 0.05", "p < 0.01")
```

```
funnel.meta(tf,
            xlim = c(-1.5, 2), contour = contour,
            col.contour = col.contour)
legend(x = 1.1, y = 0.01,
       legend = ld, fill = col.contour)
title("Funnel Plot (Trim & Fill Method)")
funnel.meta(tf.no.out,
            xlim = c(-1.5, 2), contour = contour,
            col.contour = col.contour)
legend(x = 1.1, y = 0.01,
       legend = ld, fill = col.contour)
title("Funnel Plot (Trim & Fill Method) - Outliers Removed")
```