Factors associated with mathematics attainment and educational aspirations: A comparative psychometric study of Japan, Türkiye, and England

Tevfik Can Karabıyık

Submitted in accordance with the requirements for the degree

of

Doctor of Philosophy

The University of Leeds

School of Education

March 2023

Authors Declaration

Statement 1

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Statement 2

Acknowledgements

First and foremost, I would like to sincerely thank my supervisors, Matt Homer and Innocent Tasara, for their valuable guidance, continuous support, and patience during my doctoral studies. Their invaluable advice and constructive criticism during my research and writing process were greatly appreciated and helped me to stay on track. I am grateful for their unwavering support throughout my doctoral studies. It has always been a privilege for me to have supervisors like you.

I wish to express my sincere thanks to the Ministry of National Education of Türkiye for awarding me a scholarship to study abroad, allowing me to complete my master's and doctoral studies in the United Kingdom. I am extremely grateful for the opportunity to gain valuable academic experience and knowledge abroad provided by the Ministry of National Education of Türkiye.

Furthermore, I would like to extend my deepest gratitude to my beloved wife, Elif, for her unconditional support, love, and tolerance throughout this entire process. This could not have been achieved without your help. Last but not least, I would like to thank my parents, Mehmet and Zehra, for their encouragement and unconditional support over the past years.

Abstract

It is widely recognized that mathematics plays a fundamental role in the advancement of science and technology. There is, therefore, considerable interest in research on factors that affect mathematics achievement. The motivation of students is considered one of the most influential factors in their performance in mathematics. The expectancy-value theory (EVT) is one of the dominant theories of motivation.

This study uses Trends in International Mathematics and Science Study (TIMSS) 2019 data for 8th-grade students in Japan, Türkiye, and England. This study aims to contribute to the field in terms of methodological and substantive aspects. Methodologically, it examines the psychometric properties (factor structure, method effects of negative coded items, reliability) of TIMSS 2019 motivation measures of Japan, Türkiye, and England. It also compares confirmatory factor analysis (CFA) and exploratory structural equation modelling (ESEM) measurement models in terms of model fit. Substantively, it investigates the relationship between students' motivation towards mathematics (mathematics self-concept (MSC), mathematics intrinsic value (MIV) and mathematics utility value (MUV)) and student background factors (gender and home educational resources) in predicting mathematics achievement and educational aspiration. It also examines the interaction between MSC and task value (MIV and MUV). A variety of advanced quantitative methods are used to analyse data, including factor analysis, confirmatory factor analysis (CFA), exploratory structural equation modelling (ESEM), and mediation analysis under structural equation modelling (SEM).

Methodologically, the results of the study show that the TIMSS 2019 motivation measures provide valid and reliable constructs for Japan, Türkiye, and England.

TIMSS 2019 motivation constructs support the a priori factor structure that they are designed to measure. However, negatively coded items negatively affect model fit. Furthermore, the results of this study found a high correlation between mathematics intrinsic value and mathematics self-concept. The ESEM measurement model provides a better model fit in comparison to the CFA measurement model. Substantive results show that mathematical self-concept is the variable with the strongest effect on mathematics achievement among motivational constructs. The interaction of MSC and task value also has a significant effect in predicting mathematics achievement and educational aspiration. Gender has little effect on mathematics achievement, but male students have higher MSC values than female students. Female students have higher educational aspirations than male students. Furthermore, the mediation model provides evidence that motivation constructs, particularly MSC, play a significant mediating role in the relationship between gender and HER and mathematics achievement.

Overall, this study makes a key contribution to analysing the psychometric properties of TIMSS 2019 motivation data and understanding its relationship with mathematics achievement and educational aspirations. These findings should be used to inform the development of TIMSS motivation measures and the future development of educational practitioners' understanding of the relationship between student motivation and mathematics achievement.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

ASC	Academic self-concept
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CTCU	Correlated traits, correlated uniqueness
DfE	Department for Education, England
EFA	Exploratory factor analysis
ES	Expectancy for success
ESEM	Exploratory structural equation model
EVT	Expectancy-value theory
ExV	Expectancy value interaction
FIML	Full information maximum likelihood
GVNQ	General national vocational qualification
HER	Home educational resources
HOUWGT	Student house weight
IEA	International Association for the Evaluation of Educational Achievement
ILSA	International large-scale assessments
IRT	Item Response Theory
LMS	Latent moderated structural equations
MEXT	Ministry of Education, Culture, Sports, Science and Technology, Japan
MIMIC	Multiple-indicator-multiple-indicator cause
MIV	Mathematics intrinsic value
MLR	Multiple linear regression
MoNE	Ministry of National Education, Türkiye
MSC	Mathematics self-concept

MTMM	Multitrait-multimethod
MUV	Mathematics utility value
NCETM	National Centre for Excellence in the Teaching of Mathematics, Japan
NIER	National Institute for Educational Policy Research, Japan
NVQ	National vocational qualification
OECD	Organisation for Economic Cooperation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
RMSEA	Root mean square error of approximation
SDT	Self-determination theory
SEEQ	Students' Evaluations of Educational Quality
SEM	Structural equation model
SENWGT	Student senate weight
SES	Socioeconomic status
STEM	Science, Technology, Engineering, and Mathematics
TIMSS	Trends in International Mathematics and Science Study
TLI	Tucker-Lewis index
TOTWGT	Total student weight

Chapter 1: Introduction

1.1. Introduction

Education plays an important role in establishing a country's competitive advantage in today's growing global economy. Mathematics education is of great national importance, being at the forefront of science and technology in this technologically oriented era (Ker, 2013). As a result of globalisation and the rapid development of new technologies, educational policies need to respond to these changes. The development of high technology requires a solid understanding of mathematics. It has been identified that students' interest and motivation to learn mathematics are crucial during the secondary school years (Brophy et al., 2008; Dawes and Rasmussen, 2007).

Large-scale international assessments provide ample data on education-related policies and practices to monitor education systems and make global comparisons. The Trends in International Mathematics and Science Study (TIMSS) is one of the most respected among large-scale international assessments. This study assesses international students' achievement in mathematics and science at the fourth- and eighth-grade levels. The most recent assessment, TIMSS 2019, included students from 64 countries in the fourth (average age 10) and eighth grades (average age 14) (Mullis et al., 2017).

This thesis examines the psychometric properties of the constructs measuring students' motivation towards mathematics in TIMSS 2019 and investigates the relationship between motivation and mathematics achievement and educational aspirations. Specifically, in TIMSS 2019, students' attitudes towards mathematics were measured through the constructs "students like mathematics", "student value mathematics", and "student confidence in mathematics". Firstly, this thesis comparatively investigates the psychometric properties of these constructs provided by TIMSS in countries with different demographic and cultural backgrounds, such as Japan, Turkiye, and England. It then analyses the relationship between motivational factors, gender, home educational resources (HER) and mathematics achievement and educational aspiration within the framework of expectancy-value theory (Eccles, 1983).

This chapter presents an overall view of the study. First, this chapter starts with descriptive information about TIMSS, as the data source of this thesis. Then, the chapter examines the education systems and mathematics curricula of the countries involved in the study (Japan, Türkiye, and England) and the impact of TIMSS on their educational systems. Next, we provide a brief overview about the key theoretical concepts underpinning in the study, followed by a brief mention of the problem statement, the purpose of the study, and its importance. Finally, this chapter ends with a section describing the thesis's overall structure and the organisation of the other chapters.

1.2. The Data Source – TIMSS

International large-scale assessments (ILSAs) provide information about a particular curriculum area and overall levels of student achievement in an educational system. The importance of ILSAs has continued to increase over the last two decades as a means of monitoring and improving the quality of education systems (Clarke and Luna-Bazaldua, 2021). Furthermore, ILSAs have been used to monitor whether student achievement standards have been rising or falling over time (Greaney and Kellaghan, 2008). These assessments not only measure mathematics or science achievement but also collect data from teachers, school principals, parents, and students. These include educational contexts such as gender performance, students' background, home environment, students' attitudes towards learning, school facilities, educational support, availability of resources, curriculum and teaching approaches, and teacher preparation in teaching. This serves to facilitate identification of the factors that directly affect achievement. In addition, countries can monitor their education systems and analyse those of successful countries from a comparative perspective through the data provided by ILSAs.

The most well-known ILSAs are the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), organised by the International Association for the Evaluation of Educational Achievement (IEA), and the Programme for International Student Assessment (PISA), organised by the Organisation for Economic Cooperation and Development (OECD). There are important differences between PISA, TIMSS, and PIRLS. The latter two (TIMSS and PIRLS) are curriculum-based and require certain content coverage, while PISA is focused on students' ability to apply their skills in the modern world (Fishbein et al., 2021; OECD, 2019). While TIMSS measures the performance of the fourth (aged 9–10) and eighth (aged 13–14) grades, PISA investigates the performance of the ninth grade (aged 15) (Fishbein et al., 2021; OECD, 2019). The latter provides background data related to reading due to its focus on literacy in the latest PISA (2018); TIMSS, on the other hand, provides mathematics-related background data, such as information on school resources, student attitudes, instructional practices, and home support. The data in this thesis are derived from the most recent wave of the TIMSS from 2019 (Fishbein et al., 2021) because of my personal interest in mathematics achievement and the related factors that go with it. A more detailed description of the data set can be found in Section 3.2 of Chapter 3.

The TIMSS programme includes an assessment of the mathematics and science knowledge of fourth (or fifth) and eighth (or ninth) grade students in various countries worldwide. The International Association for the Evaluation of Educational Achievement (IEA) conducts the TIMSS survey, which allows participating countries to compare cross-border educational achievements in mathematics and science and monitor educational achievement progress over time. Contextual information about students' homes, schools, and classrooms is also collected by TIMSS to explain their academic achievement. The assessment consists of projects carried out on a four-yearly basis since 1995, across grade levels four and eight. The latest cycle of TIMSS was conducted in 64 countries (Mullis et al., 2020) in 2019.

The TIMSS mathematics and science scale was developed to range from 0 to 1,000 by taking the distribution of achievements of the participating countries as a reference. But in practice, student performances usually range between 300 and 700 (Mullis et al., 2020). The mean of the overall distribution of achievement is 500 points, which is the centre point of the scale, while the standard deviation of distribution is 100 points. Assessment data from subsequent cycles have been linked to these scales to determine whether average achievement has increased or decreased over time.

Figure 1. 1 International benchmarks of mathematics achievement in TIMSS 2019

	Advanced International Benchmark
625	*Students can apply and reason in a variety of problem situations, solve linear equations, and make generalizations.* They can solve a variety of fraction, proportion, and percent problems and justify their conclusions. They can understand linear functions and algebraic expressions. Students can use their knowledge of geometric figures to solve a wide range of problems involving angles, area, and surface area. They can calculate means and medians, and understand how changing data points can impact the mean. Students can interpret a wide variety of data displays to draw and justify conclusions, and solve multistep problems. They can solve problems involving expected values.
	High International Benchmark
550	*Students can apply their understanding and knowledge in a variety of relatively complex situations.* They can solve problems with fractions, decimals, ratios, and proportions. Students at this level show basic procedural knowledge related to algebraic expressions and equations. They can solve a variety of problems with angles, including problems involving triangles, parallel lines, rectangles, and congruent and similar figures. Students can interpret data in a variety of graphs and solve simple problems involving outcomes and probabilities.
	Intermediate International Benchmark
475	*Students can apply basic mathematical knowledge in a variety of situations.* They can solve problems involving whole numbers, negative numbers, fractions, decimals, and ratios. Students have some basic knowledge about properties of two-dimensional shapes. They can read and interpret data in graphs and have some rudimentary knowledge of probability.
	Low International Benchmark
400	*Students have some knowledge of whole numbers and basic graphs.*

Source: IEA's Trends in International Mathematics and Science Study TIMSS 2019 Downloaded from http://timss2019.org/download

The TIMSS international benchmarks levels in mathematics are summarised in Figure 1.1. There are four levels on the TIMSS eighth-grade mathematics achievement scale about students' performance on the assessment items. TIMSS describes achievement at each point along this scale as an international benchmark that allows for interpretation of the results. These are the "Advanced International Benchmark (625)", "the High International Benchmark (550)", "the Intermediate International Benchmark (475)", and "the Low International Benchmark (400)" (Fishbein et al., 2021). For example, at the Low International Benchmark, students demonstrate knowledge of whole numbers and basic graphs while at the Advanced International Benchmark, they demonstrate some knowledge of complex situations and reasoning.

The TIMSS 2019 mathematics assessment is known to encompass two different aspects for students in Grade 8. The first of these centres on content, which spans four different fields, namely algebra, data and change, geometry, and number; the second pertains to cognition, which comprises knowing, applying, and reasoning; the knowing area evaluates the knowledge of the learner in regard to mathematical factors, concepts, and procedures; whereas the applying area considers the degree to which students are able to apply their knowledge to overcome problems. Reasoning is centred on the overall ability of the student to overcome complex contexts, unfamiliar situations, and multi-step problems (Grønmo et al., 2015). Figure 1.2 shows the target percentages for content and cognitive domains.

Figure 1. 2 Content and cognitive assessment target percentages

Content Domain	Percentage
Number	30%
Algebra	30%
Geometry	20%
Data and Probability	20%

Cognitive Domain	Percentage
Knowing	35%
Applying	40%
Reasoning	25%

Source: IEA's Trends in International Mathematics and Science Study TIMSS 2019 (Mullis et al., 2019)

This thesis utilises TIMSS 2019 data from Japan, Türkiye, and England. In the following section, the reasons for selecting these countries and information about their education systems are given.

1.3. Country Educational Profiles

The purpose of this section is to provide relevant information on the education systems of England, Japan, and Türkiye. Therefore, as it is a comparative study in nature, it is important to understand the educational systems of these countries and the different impacts of TIMSS in these countries.

Two significant reasons have led to the selection of Japan, Türkiye, and England for this study; first of all, Japan is one of the countries with the highest TIMSS scores in mathematics; with a total of 594 mathematics achievement score, Japan had one of the best performance in this category. England's performance was above the central point (500) with a mathematics achievement score of 515, whereas Türkiye was slightly below the central score (500) with a score of 496. This selection thus allows the investigation of the impact of student motivation in countries with different levels of mathematics achievement. The second reason was to examine the critique that TIMSS motivation measures are based on the West and that the factor structures, validity, and the concern that the reliability of these scales might be lower in countries other than Western countries (Bofah and Hannula, 2015; Marsh et al., 2013a; Metsämuuronen, 2012; Rutkowski and Rutkowski, 2010). Therefore, England is selected as a Western country; as an Asian country Japan represents a different culture, and Türkiye is a culturally diverse country, arguably with characteristics of both the West and the East.

For each country, I first explain the educational structures of the countries and then explain how TIMSS has affected their education systems and educational policy and reforms. Then I present the countries' past TIMSS performance, and finally, I give an overview and objectives of the mathematics curriculum in each country. This will help me to cross-culturally discuss the psychometric properties of the TIMSS motivation measures and their relationship with mathematics achievement and educational aspiration (see Discussion in Chapter 7).

1.3.1. Education system of England

The Department for Education (DfE) is responsible for the school system in England (DfE, 2019). A total of 8.82 million students attended 24,323 schools and 257 further education colleges in 2019 (Isaacs et al., 2020). Full-time education is compulsory for children between the ages of 5 and 16 years old. The legal requirement for young people is to enrol in a full-time educational programme, an apprenticeship or traineeship, or a part-time educational programme along with paid or volunteer work until the age of 18 (Greaney and Kellaghan, 2008).

Primary school students typically move to secondary school at 11 years old. In most secondary schools, students can remain in education until the age of 18; however, students may also attend Sixth Form colleges, FE colleges, apprenticeships, or traineeships as early as 16 years old (Isaacs et al., 2020). A state-funded school can be classified as a maintained school by a local authority, an aided school by volunteering, an academy, or a free school. “All schools, state-funded or independent, are required to provide a broad and balanced curriculum, and there are statutory requirements for particular subjects” (Isaac et al., 2020, p.1).

The country introduced a new national curriculum between 2014 and 2016, which is available to academies, free schools, and independent schools (DfE, 2014). Every key stage has its own set of programmes of study that define performance expectations (Isaac et al., 2020).

1.3.1.1. The mathematics curriculum in England

Students taking TIMSS assessments in Grade 8 (Year 9) are generally expected to have completed most of the Key Stage 3 mathematics programme of study, which is designed to ensure that all students are proficient in the following areas (Isaacs et al., 2020):

- Students need to develop conceptual understanding and be able to recall and apply information quickly and accurately in order to become fluent in the fundamentals of mathematics.
- Using mathematical language to develop an argument, justification, or proof based on a line of inquiry, assuming relationships, and generalising.
- Utilise mathematics to solve problems of increasing complexity, following simple steps and persevering to find solutions.

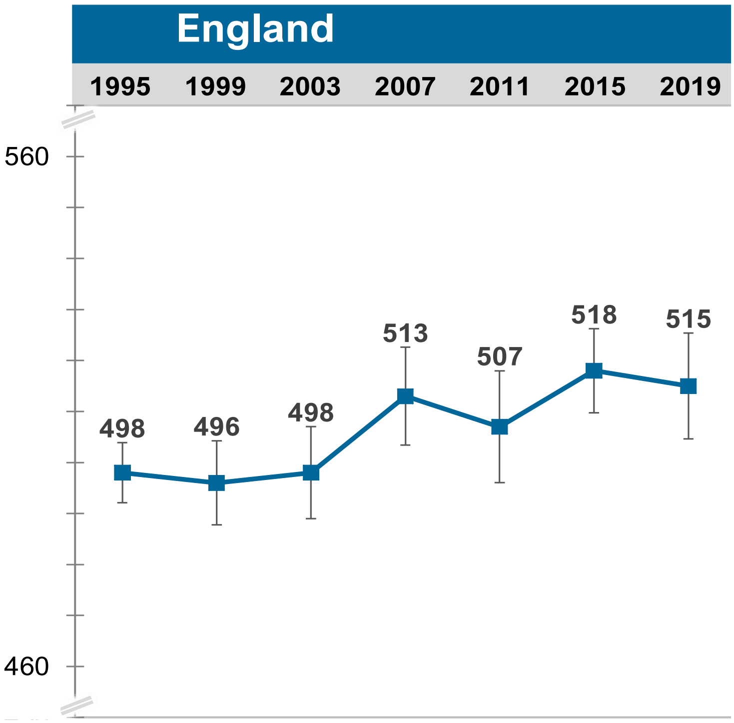
1.3.1.2. Impact of TIMSS in England

England has participated in all TIMSS cycles since 1995. Figure 1.3 illustrates the trend in eighth-grade students' mathematics scores in England. The table shows that English eighth-grade students achieved slightly above the centre point of 500 points after 2007.

In high-performing countries such as Japan, a higher percentage of students achieved Advanced and High TIMSS International Benchmarks in both subjects; but in England, the achievement gap between more advantaged and less advantaged students was larger than in most other countries with high academic achievement (Richardson et al., 2020). This means that the achievement gap between students is less than in England for top achievers in countries such as Japan. In other words, the majority of students perform well in countries with high performance in TIMSS.

In England, participating schools are provided with confidential feedback to support their professional development and school improvement (Isaacs et al., 2020). In addition, TIMSS school conferences are held to discuss national results and share improvement ideas. The results of TIMSS have been used to determine improvement priorities for policy and practice (in conjunction with other international studies). These priorities can be identified through the National Centre for Excellence in the Teaching of Mathematics (NCETM) and the National Science, Technology, Engineering, and Mathematics (STEM) Learning Centres (Isaacs et al., 2020).

Figure 1. 3 Trends in mathematics in England at Grade 8 (Year 9)



Source: IEA's Trends in International Mathematics and Science Study TIMSS 2019 (Mullis et al., 2019)

1.3.2. Education system of Japan

The fundamental law of education in Japan was legislated in 1947 and revised in 2006 (Hou, 2006). The law outlines the fundamental principles of Japanese education and ensures that all students have access to free, compulsory education for nine years

(Kelly et al., 2020). School education is administered by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Almost all public schools are established and maintained by local governments, which are accountable to prefectural or municipal education boards.

At every level of the academic hierarchy, there are both public and private institutions. National government is responsible for covering the majority of the expenses associated with national schools, while municipalities and prefectures are responsible for supporting their own schools with some government assistance (Kelly et al., 2020). The Japanese educational system has three levels of education: six years of primary school, three years of lower secondary school, and three years of upper secondary school. Students may also attend six-year secondary schools with a combination of lower and upper secondary education (MEXT, 2019).

Almost all children aged 6–15 are enrolled in school, which is compulsory and requires them to complete six years of primary and three years of lower secondary education. The percentage of students enrolled in upper secondary school in 2018 was 98.8%, while for those entering higher education it was 53.3% (MEXT, 2019). A full-time, part-time, or correspondence education is available in upper secondary schools. Most students finish upper secondary school in three years in full-time schools, but it may take longer in part-time and correspondence schools. The percentage of upper secondary students enrolled full-time in 2018 was approximately 97.3% (MEXT, 2019).

1.3.2.1. The mathematics curriculum in Japan

It is a requirement for students in primary, lower secondary, and upper secondary schools to study mathematics. Mathematics activities have been included in the curriculum objectives since the 1998 revision of the mathematics curriculum (Kelly et al., 2020). In addition, students at the primary and lower secondary levels are expected to enjoy mathematics, while students at the upper secondary level are expected to develop creativity in mathematics as an objective (Kelly et al., 2020). Curricula for mathematics are structured in three parts: a general objective, a grade-specific objective, and a syllabus. In addition, objectives and content, teaching plans, and remarks on content specify methods and materials to some extent. It is furthermore a requirement of the primary school curriculum that mathematics be covered in a certain number of class periods each year. Providing quality instruction in mathematics content is a requirement for all schools. It is the responsibility of each school to develop a mathematics strategy that includes a description of goals, content, qualities, and attitudes that should be fostered, learning activities, teaching procedures, teaching structure, and assessment strategies (Kelly et al., 2020).

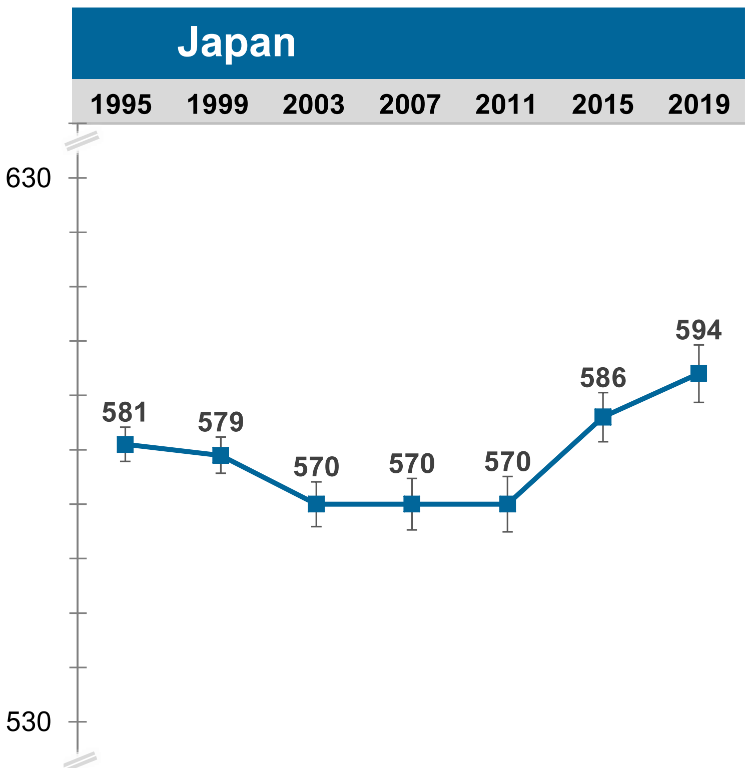
In lower secondary mathematics (Grades 7 through 9), the following objectives are intended: Enhance students' understanding of number, quantity, and geometrical concepts, principles, and rules; enhance students' ability to analyse and represent phenomena mathematically by teaching them mathematical processing and representation; promote students to use their mathematical understandings and abilities when thinking and evaluating, and encourage them to enjoy mathematics and appreciate its value (MoE, 2008).

1.3.2.2. Impact of TIMSS in Japan

TIMSS has been conducted in Japan every four years since 1995 (1999, 2003, 2007, 2011, 2015, and 2019). Figure 1.4 shows the TIMSS mathematics performance of Japanese students since 1995. As can be seen, Japanese students performed above the average score of 500 in all TIMSS. This makes it one of the top countries among those participating, with the highest achievement in mathematics.

Both PISA and TIMSS are activities which are conducted by the National Institute for Educational Policy Research (NIER) under the auspices of MEXT's educational policies, and both tests are administered by the NIER (Kelly et al., 2020). The TIMSS results have been used in several research studies addressing improvements in teaching and learning (Saruta, 2010, cited in TIMSS team in Japan, 2020). In addition, a variety of education reform discussions have referred to TIMSS results as reference material (TIMSS team in Japan, 2020). The NIER is the home of the TIMSS National Study Center, which facilitates the dissemination of TIMSS's findings to policymakers. These are particularly useful to mathematics and science curriculum specialists at NIER.

Figure 1. 4 Trends in mathematics in Japan at Grade 8



Source: IEA's Trends in International Mathematics and Science Study TIMSS 2019 (Mullis et al., 2019)

1.3.3. Education system of Türkiye

Education and training facilities are monitored, inspected, and assessed by the Ministry of National Education (MoNE, 2017). The Ministry of National Education has the following responsibilities according to the Degree Law Regarding the Organizational Chart and Responsibilities of the Ministry of National Education (MoNE, 2011):

Provide students with a variety of abilities, including physical, mental, moral, spiritual, social, and cultural skills at pre-primary, primary, and secondary levels.

- Plan, implement, and update education curricula and provide educational services for students and teachers.
- Develop, implement, monitor, and update a national policy for every level of education as necessary.
- Design an educational system that promotes innovation, is dynamic, and meets the needs of economic and social progress.
- Aim to ensure that all citizens have equal access to education.
- Ensure equitable access to education for female students as well as those with special needs.
- Develop and implement a curriculum for gifted students.

According to Turkish legislation enacted in 2012, primary and secondary education is compulsory for 12 years. It is known as the “4 + 4 + 4 Education System”, which consists of four years of primary school, four years of lower secondary school, and four years of upper secondary school (Parlak et al., 2019). The Turkish educational system is structured as follows: preschool education, basic education (which comprises primary and secondary school), secondary education (high school), and higher education.

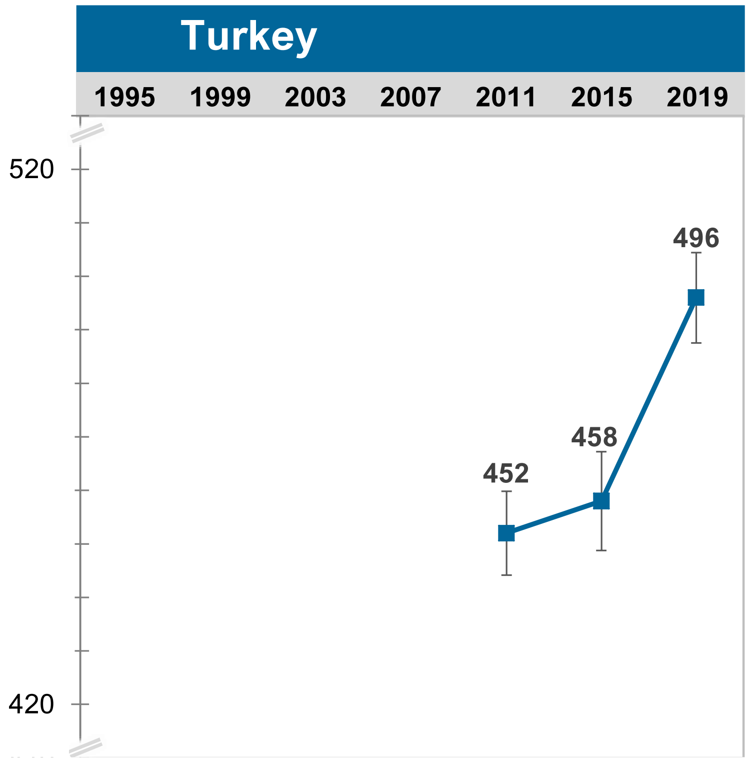
1.3.3.1. The mathematics curriculum in Türkiye

Mathematics is viewed as a tool for solving problems and sharing ideas and solutions in the current primary mathematics curriculum. Hence, the mathematics curriculum at the primary level aims to develop students' mathematical literacy and cognitive abilities, to teach them to comprehend and apply mathematical concepts in daily life (Parlak et al., 2019). The topics of the curriculum are also chosen based on the developmental level of each grade's students. Students are expected to become more active participants in mathematics through these strategies (TTKB, 2018). As part of the curriculum, students are encouraged to express their thinking and reasoning, share and discuss their solutions, and develop their research, production, and use skills (TTKB, 2018). It is anticipated that these processes will improve students' responsible and systematic characteristics. There is an emphasis throughout the primary school mathematics curriculum on conceptual and procedural understandings of mathematical concepts (TTKB, 2018). The philosophy of the education programme is to develop a positive attitude towards mathematics, as well as self-confidence in solving mathematical problems. It is intended that mathematics be integrated with other subjects and disciplines throughout the curriculum (Argün et al., 2010).

1.3.3.2. Impact of TIMSS in Türkiye

In the years 1999 and 2007 TIMSS assessments were administered to eighth-grade Turkish students, while in 2011, 2015, and 2019 they were administered to Turkish students in the fourth and eighth grades. Figure 1.5 shows the performance of Turkish students in TIMSS mathematics since 2011 (the figure taken from TIMSS and scores before 2011 are not shown in the report). The table shows that the mathematics performance of Turkish students has been on an upward trend since 2011. In the last cycle, TIMSS 2019, Turkish students scored 496 points, very close to the centre point of 500 points. This assessment is one of the international indicators used to monitor education in Türkiye and the results for students in Türkiye are considered an important reflection of the quality of their education (Parlak et al., 2019). This indirectly influences curriculum development and education reform through TIMSS and other international research projects (Parlak et al., 2019).

Figure 1. 5 Trends in mathematics in Türkiye at Grade 8



Source: IEA's Trends in International Mathematics and Science Study TIMSS 2019 (Mullis et al., 2019)

1.4. Overview of Key Theoretical Concepts

A summary of the basic theoretical concepts used in this study is presented here briefly as they are referred to in the following sections: "problem statement", "study aims", and "significance of the study". I give an overview here of theoretical concepts, with full details provided in the literature chapter.

The term “construct” refers to a phenomenon that has a theoretical basis (Edwards and Bagozzi, 2000). Alternatively, this can be used synonymously with the term “concept” (Carmines and Zeller, 1979). There are two types of constructs to consider: observed and latent variables. In statistics, latent variables can be defined as those which are not directly observable or measured, but can only be inferred indirectly through a mathematical model from other variables that are directly observable or measured (Dodge et al., 2003).

EVT is one of the theoretical approaches frequently used in broader literature to study motivational structures. It has two main components; "expectancy of success" and "task values" (for more details, see the chapter 2 section 2.3.1.3). This study examines three constructs related to expectancy-value theory (Eccles et al., 1983); “expectancy of success for mathematics”, “mathematics intrinsic value”, and “mathematics utility value”, all of which are considered to be latent variables.

Expectancy of success

Expectancy of success is measured in TIMSS with the “student confident in mathematics” scale (with items such as “I usually do well in mathematics”; “mathematics is more difficult for me than for many of my classmates”). Although TIMSS does not provide a clear theoretical basis for the development of these constructs, there is a reference to the Marsh and Craven (2006) self-concept study in the description section of the "student confident in mathematics" scale in the TIMSS assessment framework (Hooper et al., 2017).

Expectancy of success is a construct that overlaps with self-concept from a theoretical perspective (Eccles and Wigfield, 2002). It is, therefore, not surprising that several studies on expectancy-value theory (EVT) rely on academic self-concept (ASC) as a measurement of expectations of success (e.g., Guo et al., 2015; Marsh et al., 2013a; Musu-Gillette et al., 2015). Therefore, in this study, the construct measured by TIMSS as "student confident in mathematics" and conceptualised as "expectancy of success" in EVT will be accepted as a mathematics self-concept (MSC) and expressed as such.

Mathematics Intrinsic and Utility Value

Intrinsic and utility values are measured with “students like learning mathematics” and “students value mathematics” scales, respectively, in the TIMSS 2019 students’ attitudes towards mathematics concepts (see table 3.2 in section 3.2.4.1). Intrinsic value in the EVT framework is a similar concept to intrinsic motivation in self-determination theory (SDT) (Deci and Ryan, 2012). Intrinsic value and intrinsic motivation are used interchangeably in the wider literature (e.g.,Brown and Putwain, 2022; Fadda et al., 2020; Guo et al., 2015; Musu-Gillette et al., 2015; Simpkins et al., 2006). Therefore, the study uses intrinsic value and intrinsic motivation synonymously. Likewise, utility value and extrinsic motivation refer to similar concepts within EVT and SDT. Thus, similar to the operationalisation of self-concept and intrinsic value, utility value and extrinsic value are used as synonymous constructs, as can be seen in the literature (e.g.,Brown and Putwain, 2022; Fadda et al., 2020; Guo et al., 2015; Musu-Gillette et al., 2015; Simpkins et al., 2006). Table 1.1 shows the names of these constructs used in TIMSS, the concepts they are usually adapted from in the literature as synonyms, and their conceptual equivalents in this thesis. For clarity, I will use the terms mathematics self-concept (MSC), mathematics intrinsic value (MIV) and mathematics utility value (MUV) throughout this work.

Table 1. 1 Motivational constructs used in TIMSS 2019 and their synonyms

TIMSS 2019 motivational measures	**Synonyms**	**Operationalization in this thesis**
Student confident in mathematics	Mathematics Self-Concept Mathematics Self-Efficacy Expectancy of success in mathematics	Mathematics Self-Concept (MSC)
Student like learning mathematics	Intrinsic Motivation Intrinsic Value	Mathematics Intrinsic Value (MIV)
Student value mathematics	Extrinsic Motivation Utility Value	Mathematics Utility Value (MUV)

1.5. Problem Statement

This study addresses two main issues: First, methodological, and psychometric issues; there are potential problems emphasised in the literature regarding the validity and reliability of the TIMSS 2019 motivational constructs in different educational systems. This study addresses the fact that the exploratory structural equation model (ESEM), which has recently been used to investigate psychometric characteristics and develop the structural model, provides a better statistical and theoretical model than the classical confirmatory factor analysis (CFA). The second problem is substantive issues. It has been shown in numerous studies that motivation and achievement are closely related (Martin, 2001; Martin et al., 2001a;2001b; Meece et al., 1990). However, there is not only a direct relationship between achievement and motivation but also an interaction between motivational constructs (Brown and Putwain, 2022; Guo et al., 2015; Marsh et al., 2004; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012) and demographic factors – gender and socioeconomic status (SES) – that contributes significantly to the relationship (Eccles and Wigfield, 2020). This section begins with a brief overview of the potential psychometric and methodological problems in TIMSS, followed by an overview of the substantive issues associated with motivation and achievement in mathematics.

Psychometric and methodological issues. The TIMSS context questionnaire contains three scales that assess students' attitudes towards mathematics: students like learning mathematics, students' self-confidence in mathematics, and students value mathematics (Fishbein et al., 2021). However, few studies have investigated the psychometric properties (i.e., factor structure, measurement invariance) of the TIMSS background survey, even though the structural construct validity of scales in the most recent survey, TIMSS 2019, has been tested with principal components analysis by the TIMSS assessment team (Yin and Fishbein, 2019). Since principal components are not intended to analyse latent variables (Fabrigar et al., 1999), alternative validity tests, such as CFA or structural equation modelling (SEM), must be used to assess whether items are measuring the factors intended (Marsh et al., 2013a). In addition, mathematics attitude instruments have been critiqued for their weakness in justifying validity, particularly in non-Western countries because, according to some literature, TIMSS motivational questionnaires were designed in a "Western context" in terms of the psychometric elements included therein (Abu-Hilal and Aalhussain, 1997; Bofah and Hannula, 2015; Marsh et al., 2013a; Metsämuuronen, 2012; Rutkowski and Rutkowski, 2010; Wang and Berlin, 2010). In particular, the literature emphasises the importance of investigating mathematics achievement factors, as well as their psychometric properties. Furthermore, many large-scale studies are based on theories derived from Western cultures without considering that the concept of "the self" is culturally contingent and thus varies from culture to culture (Shen and Tam, 2008), which poses a challenge to theoretical application. Consequently, it has been suggested that the results garnered through the student questionnaire may not be as reliable as those collected by other countries, including Africa, East Asia, Europe (in parts), and the Middle East, when contrasted alongside North America, where the design of the scales originated.

In addition, an increasing number of studies have showed that the ESEM is more appropriate and provides a better model parameters in latent variable modelling than CFA (Furnham et al., 2013; Guay et al., 2015; Marsh et al., 2010; Marsh, Herbert W et al., 2014; Marsh et al., 2009; Marsh et al., 2013c; Morin et al., 2013). To the best of my knowledge, comparative CFA and ESEM model fits of different countries using TIMSS 2019 mathematics motivation data have not yet been conducted in the literature. In this regard, one of the objectives of this research is to contribute to the existing literature to fill the gap in the evaluation of psychometric properties of TIMSS 2019 motivation measures, and the comparison of CFA vs ESEM approaches.

Substantive issues. Knowledge and skills relating to mathematics are critical to economic societies focused on science and technology (Ker, 2017). Nonetheless, the percentage of learners opting to become involved in science- or mathematics-related fields has dropped worldwide (Dawes and Rasmussen, 2007). In secondary schools, academic achievement is the key focus when it comes to pursuing improvements, and so the initial stage in this process is to encourage learners to become involved in science- and mathematics-related learning tasks while simultaneously pursuing careers or majors in this field. It is essential that the link between any changes in significant motivational factors of the career path and involvement in maths-related subject areas in teaching is well understood, as highlighted by, for example, Nagengast et al. (2011). Moreover, research concerned with motivational and engagement change can facilitate educators in providing suitable designs for motivational interventions at different points of the learning process and educational career. Furthermore, motivation provides a key foundation for achievement (Martin, 2001; Martin et al., 2001a;2001b; Meece et al., 1990). In this regard, it is important to identify the reasons that affect mathematics attainment in order to increase mathematics achievement. The expectancy-value theory, initially developed by Atkinson (1957), has been one of the influential theories in understanding the complex relationship between achievement and motivation for decades (Eccles, 1994;2009). This theory proposes that performance and choices related to achievements are determined by the expectancy of success and the level of value attached to a given task (Eccles, 1994;2009). The work of Eccles and colleagues (Eccles et al., 1983) examines subjective task values and motivational beliefs in relation to other psychological, social, and cultural factors. However, the interaction relationship of expectancy and values, which is essential to classical EVT (Atkinson, 1957), has received less scrutiny in modern EVT (Nagengast et al., 2011; Trautwein et al., 2012). The reason for this gap might be the lack of advanced statistical techniques that are suitable for measuring the interactions between expectancy and value in non-experimental study. (Guo et al., 2015). The recent development of latent variable approaches to interaction effects has made it possible to analyse more accurately the latent interactions inherent in classical EVT (Brown and Putwain, 2022; Guo et al., 2015; Marsh et al., 2004; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012). However, with the exception of Guo et al. (2015), these empirical studies have examined the interaction relationship by considering only one component of task values with expectancy, but this is inconsistent with the assumption that multiple task values of EVT simultaneously influence outcomes related to achievement. In the EVT model, expectation and task values are assumed to mediate the relationship between children’s backgrounds, including gender and family

SES and educational outcomes. In recent studies, motivational beliefs have been examined as a mediator of the relationship between SES, gender, and educational outcomes (de la Fuente et al., 2013; Nagy et al., 2008; Nagy et al., 2006; Parker et al., 2012). However, few studies have compared the direct and indirect effects of motivational beliefs when considering both expectation and task values simultaneously (Guo et al., 2015). In summary, the critical issue here is that the relationship between SES, gender, motivational constructs, and achievement has generally been investigated through direct effects. Therefore, in this study, the mediating effect of motivational constructs on the relationship between gender and SES and achievement will be highlighted.

1.6. Study Aims

This study aims to contribute to the literature in five overarching aspects: (1) a comparison of CFA and ESEM approaches with TIMSS 2019 motivation data; (2) an evaluation of the psychometrics properties, such as factor structure, measurement invariance, reliability, method effects of TIMSS 2019 motivational constructs by comparing Japan, Türkiye, and England; (3) an examination of the effect of EVT factors on mathematics achievement and educational aspiration across Japan, Türkiye, and England; and (4) testing the expectancy-value interaction which was present in the original EVT (Atkinson, 1957) but is not found in modern EVT (Eccles, 1983; Nagengast et al., 2011); and (5) measuring the mediating effect of EVT factors in the relationship between background factors and educational outcomes.

1.7. Significance of the Study

The present thesis is a substantive-methodological synergy (Marsh and Hau, 2007), which incorporates current methodology to address substantively significant issues related to mathematics motivation (self-concept, intrinsic value, and utility value). Using data from the international TIMSS, students from Japan, Türkiye, and England are compared to their responses to motivation constructs. Substantively, this study is based on the theory and research associated with the EVT. From a methodological standpoint, this study aims to analyse the psychometric properties of TIMSS motivational constructs, test their measurement invariance across different cultures, and conduct these psychometric analyses by comparing CFA and ESEM factor analyses. The TIMSS data has the potential to provide a robust cross-cultural perspective, so it is also important to assess whether theoretical models can be generalised across cultures (Marsh et al., 2006).

This work also investigates the relationship between SES, the attitudes of learners, and educational outcomes, not only as a direct relationship but also with interaction and mediation effects. The findings will aid understanding about the effect of such factors on academic attainment in mathematics. In addition, this work will contribute to the literature available in fields relating to learners' viewpoints about mathematics and will further assist in generalising such constructs to different contexts. This is owing to the recognition that most psychometric properties-based literature in specific consideration to motivational constructs is carried out in a Western context with unquestionably excellent psychometric properties (Heine, 2001; Markus and Kitayama, 1991; Marsh et al., 2013a; Shen and Tam, 2008). Owing to the fact that the TIMSS data adopt a cross-country perspective, it provides a valuable source of reference should future researchers seek to determine generalised theories, as studies that bring together and combine various cultural viewpoints are crucial when it comes to identifying more universal and valuable theories (Van de Vijver and Leung, 2000).

In summary, this research allows us to comparatively examine how the EVT factors that influence student achievement differ across different cultures/educational systems. This study is also intended to contribute to modern EVT by examining the interaction effect of “expectancy” and “task values” that previously existed in the original EVT but are ignored in the modern EVT. In addition, it is also important to examine the mediating effects of EVT factors in the relationship between background variables and mathematics achievement and the desire for education.

1.8. Organisation of the Thesis

This thesis consists of eight chapters. This chapter (Chapter 1) provides an overview of the study. Chapters 2 and 3 provide a context for the study along with an explanation of its rationale. The original empirical findings are presented in Chapters 4, 5, and 6. The thesis concludes with a discussion and conclusion in Chapters 7 and 8.

Chapter 2: Literature review. This chapter reviews the methodological literature on SEM, factors analysis, and the substantive work on the association between student motivation and mathematics achievement. I first outline the extant literature on the analysis of latent variables with ESEM and CFA. This is then followed by a brief explanation about commonly used theories on student motivation: self-determination theory, achievement-goal theory, self-concept, and expectancy-value theory. Later, I present empirical research on the association of students’ motivation with mathematics self-concept (MSC), mathematics intrinsic value (MIV), and mathematics utility value (MUV), and background factors – gender and home educational resources (HER) – and educational outcomes (mathematics achievement and educational aspirations). Finally, I present a summary and state the research question and conceptual framework of the thesis.

Chapter 3: Data and methods. The purpose of this chapter is to describe the data and methods used in this study. A discussion of the rationale for selecting the TIMSS 2019 data set is provided in the first section, followed by an examination of its main characteristics. It also provides an explanation of the methods that were chosen to answer the research questions, along with a justification for their selection. An explanation in relation to a number of significant decisions taken about several data-related issues and methodological considerations that had to be addressed is offered. Finally, I conclude the chapter with a brief comment on ethical considerations.

Chapter 4: Evaluation of psychometric properties of TIMSS 2019 motivation measures and their relations with demographic variables and educational outcomes.

This section starts by comparing the CFA and ESEM measurement models in terms of the model fit indices, parameters, and factor structures of motivation factors. Second, the psychometric properties of motivation measures, such as factor loadings, measurement invariance and reliability, are discussed due to the comparison of the these two models. Finally, a multiple indicators multiple causes model (MIMIC) is developed by adding educational outcome variables and demographic variables as covariates into the model to evaluate the convergent and discriminant validity of motivational factors and their relations with covariates based on correlation analysis. My objective in this chapter is to determine which statistical modelling approach is more appropriate (CFA vs ESEM) as well as evaluate the psychometric aspects of TIMSS 2019. The results of this analysis are crucial for the design of the predictive models in the chapters to come.

Chapter 5: The effects of expectancy-value beliefs and their interaction on educational outcomes.

This chapter studies the effects of motivation factors on educational outcomes (mathematics achievement and educational aspirations) within the framework of EVT with SEMs. This section also examines the latent interactions between expectancy and value in predicting educational outcomes with the unconstrained product indicator approach under the structural equation model. This chapter aims to investigate the power of predicting the educational outcomes of motivation factors within the scope of EVT in Japan, Türkiye, and England. Additionally, this chapter examines the interaction effect of expectancy (self-concept) and value beliefs (intrinsic and utility value).

Chapter 6: The mediation effect of expectancy-value components on educational outcome.

This chapter seeks to explore the relationship between HER, gender, and mathematics achievement, and educational aspirations with the mediating role of expectancy-value factors, namely MSC, MIV, and MUV. The aim of this section is to explore the effect of gender and HER on educational outcomes through student motivation.

Chapter 7: Discussion. In this chapter, the results of this thesis are discussed. I consider the findings from Chapters 4 through 6 related to the research questions and critical aspects of the literature review presented in Chapter 2. The limitations of this study are also discussed in this chapter.

Chapter 8: Conclusions. The final chapter of the thesis concludes the thesis. This chapter discusses the background and scope of the research, describes how it achieved its objectives, and summarises its results. The purpose of this chapter is to provide a brief overview of the research findings, discuss the policy, theory, and practice implications of this study, and highlight some of its primary contributions. Finally, in conclusion, some recommendations are made for future research.

Chapter 2: Literature Review

2.1. Introduction

The literature chapter consists of two main sections: "methodological background of the study" and "substantive literature on motivation".

This study focuses on methodology as one of its main research components. Therefore, the first section of the literature chapter provides an overview of the relevant methodological issues discussed in the literature. The second section reviews the theoretical and empirical literature related to the study with regard to the motivation to learn, background characteristics, mathematics achievement, and educational aspirations. The literature review in this study focuses primarily on motivation studies with large-scale assessment data, particularly TIMSS data. This decision was made in accordance with the particular focus of the study since there is extensive literature regarding student attitudes and motivation regarding mathematics in educational psychology. The use of TIMSS data in this study was chosen in order to provide a more comprehensive view of the global trends in student attitudes and motivation research and to enable an analysis of these trends across different countries. The TIMSS data set is particularly comprehensive and allows for a detailed and meaningful comparison of student attitudes and motivation across different countries and has been extensively used in academic research.

2.2. Methodological Background of the Study

An increasing number of studies show that ESEM is more appropriate than CFA in the analysis of overlapping theoretical constructs, such as motivational constructs, and provides a better measurement model fit through its flexible structure that allows cross-loading between factors (Furnham et al., 2013; Guay et al., 2015; Marsh et al., 2010; Marsh et al., 2009; Marsh et al., 2013c; Morin et al., 2013). Therefore, this section presents a literature review of studies comparing CFA and ESEM as measurement models and examines the advantages and disadvantages of each.

2.2.1. The comparison of CFA and ESEM as a measurement model

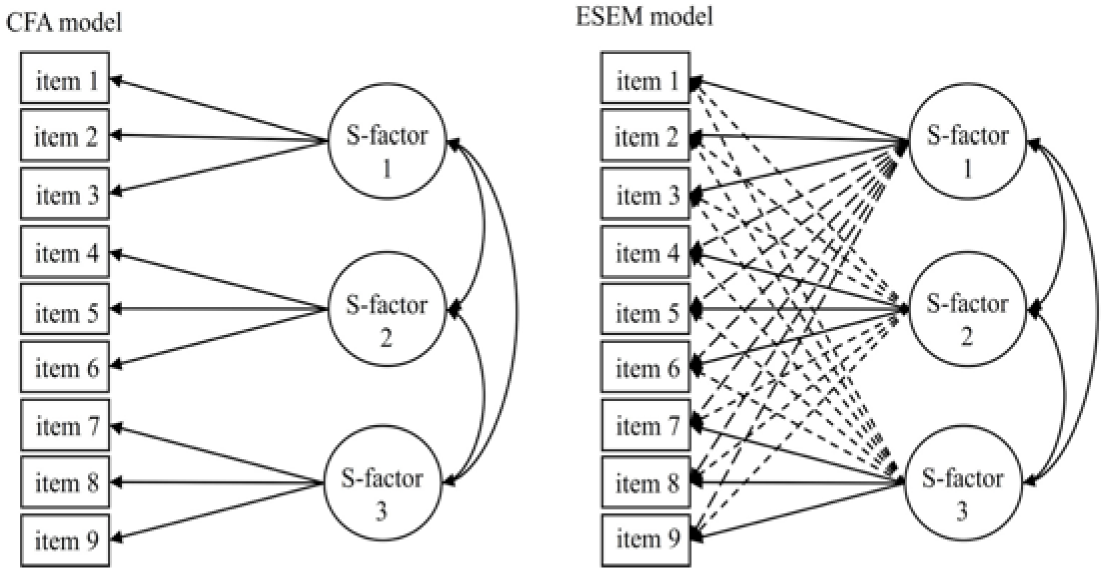
Confirmatory factor analysis has been one of the most preferred methods for the analysis of multidimensional instruments for decades (Morin et al., 2016). The CFA and SEM frameworks have significantly influenced psychological and educational research (Bollen, K. A., 1989; Jöreskog et al., 1973). The basic assumption of CFA is that the observed indicators (items) are linked to certain targeted factors based on previous theory or analysis. This procedure is consistent with the constraining independent cluster model of CFA (ICM-CFA) and allows researchers to develop more parsimonious models. However, CFA's independent cluster model requires strong measurement assumptions that do not always match the real data (Guay et al., 2015). Restricting cross-loading on untargeted factors to zero in ICM-CFA models is often overly simplistic, restrictive, and idealistic (Asparouhov and Muthén, 2009). Specifically, in social sciences, most items have many cross-loadings (although much weaker than the main-loading) conceptually related to one another (Guay et al., 2015; Marsh et al., 2010). Ideally, fitting such data to all non-targeted factors with zero cross-loading provides a parsimonious model, but may not be accurate for the data and theory (Xiao et al., 2019). These misidentified models can cause inflation for factor correlations that misinterpret the relationships between constructs and thus their meaning (Marsh et al., 2010; Marsh, Herbert W. et al., 2014; Marsh et al., 2009). Therefore, the resulting misfit models will need to be revised to reach a better model fit for the final models.

On the other hand, exploratory factor analysis (EFA) is a measurement model that provides an estimation of cross-loadings. Therefore, it is required to test conceptually related constructs for multidimensionality (Morin et al., 2016). At its core, EFA looks for possible latent variables (or factors) based on observed variables (or items) included in the data, and is thus mainly performed when researchers do not have clear information about the factor structure of their data (Alamer and Marsh, 2022). However, EFA has often been criticised for being data driven and “exploratory” in nature (Guay et al., 2015; Kahn, 2006). This refers to an approach where multiple models are compared, and the model that best fits the data (based on various criteria) is stored for later use. However, scholars argue that the validity of constructs cannot be thoroughly evaluated using EFA alone (Alamer, 2021; Hair et al., 2019; Marsh et al., 2009). This is due to the limitations of EFA compared to CFA, such as lack of fit indices and validity and measurement invariance across groups. In contrast, CFA is generally assumed to be theory driven. Models are evaluated internally using various fit indices and allow the testing of group differences by multigroup measurement invariance test. According to Morin et al. (2013, p.396),

> This perception is reinforced by the erroneous semantically based assumption that EFA is strictly an exploratory method that should only be used when the researcher has no a priori assumption regarding factor structure and that confirmatory methods are better in studies based on a priori hypotheses regarding factor structure. This assumption still serves to camouflage the fact that the critical difference between EFA and CFA is that all cross-loadings are freely estimated in EFA. Due to this free estimation of all cross-loadings, EFA is clearly more naturally suited to exploration than CFA. However, statistically, nothing precludes the use of EFA for confirmatory purposes,

except perhaps the fact that most of the advances associated with CFA/SEM were not, until recently, available with EFA.

Figure 2. 1 Simplified graphical representation of CFA, ESEM



Note: CFA, confirmatory factor analysis; ESEM, exploratory structural equation modelling; S-factor, specific factors. The full one-headed arrow represents main factor loadings, dashed one-headed arrows represent cross-loadings, and two-headed arrows represent correlations. Source (Tóth-Király et al., 2017)

To overcome the limitations of EFA and CFA, the ESEM approach has been developed by combining the advantageous aspects of these methods (Asparouhov and Muthén, 2009). In this way, the ESEM approach allows cross-loading between items and factors simultaneously, like an EFA model, while also calculating goodness-of-fit indices, allowing error terms to be estimated, and measurement invariance like a CFA model (Alamer and Marsh, 2022). In nearly all multidimensional studies, ESEM fits data better due to its flexibility and less restrictive procedure compared to CFA . Factor correlations, even if small and insignificant, tend to be unbiased and reflect data appropriately when cross-loadings are estimated (Marsh et al., 2014). For these reasons, ESEM's fit indices tend to be better than CFA (Alamer, 2022). In Figure 2.1, CFA, bifactor CFA, ESEM, and bifactor ESEM models are presented visually. The solid lines in the ESEM model represent the item loadings on the main target factor, while the dashed lines show the cross-loadings.

2.2.2. Empirical studies on the comparison of CFA and ESEM

Research based on ESEM has recently been used to evaluate multidimensional constructs in the field of psychology and education (Morin et al., 2013). The first empirical studies based on ESEM were conducted by Marsh et al. (2009), in which they examined substantively important questions in relation to the perceptions of university students with regard to the quality of the faculty's teaching using a multidimensional 36-item SEEQ (Students' Evaluations of Educational Quality) questionnaire. A priori nine-factor solutions in EFAs have largely been supported (Marsh and Hocevar, 1991), yet CFAs have not been able to replicate the results (Toland and De Ayala, 2005). The results of Marsh et al. (2009) support previous EFA research by demonstrating that ESEM structures were able to fit the data whereas ICM-CFA models could not. A critical finding was that the SEEQ factor correlations in the CFAs were inflated more than in the ESEMs, undermining their discriminant validity as diagnostic criteria when used in CFAs. It means ESEM model provide more optimistic factor correlations than corresponded CFA model.

A number of ESEM studies have tested the Big Five personality factor constructs (Furnham et al., 2013; Marsh et al., 2010; Marsh et al., 2013c). Marsh and colleagues used a new and evolving ESEM methodology for the assessment of Big Five personality responses. Marsh et al. (2010) employed ESEM in order to reduce the high correlation among latent factors in the Big Five-factor structure for responses to the 60-item NEO-FFI (Neuroticism, Extraversion, and Openness Five Factor Personality Inventory). Later, Marsh et al. (2013b) employed ESEM to simulate how Big Five components vary throughout life with age, gender, and interaction factors, using the 15-item Big Five Inventory from the British Household Panel Survey (N = 14,021; 15–99 years). As another example, Furnham et al. (2013) performed an ESEM analysis on a sample-based Big Five response (240-item NEO-PI-R) in a high-risk work-related context. The results of all of these studies agreed with Marsh et al. (2009) that ESEM provided more accurate factor correlations than CFA and fitted data significantly better than CFA. In addition, a recent study by Alamer and Marsh (2022) provides evidence of the effectiveness and flexibility of ESEM. These scholars are conducting research in the field of second language research and collected two sets of data in order to create the second language (L2) Passion Scale, which measures a dualistic model of passion for the second language. A total of 220 L2 students participated in this study and CFA and ESEM models were compared. According to the results, the ESEM method offers significantly better goodness-of-fit indices and realistic correlation factors in comparison with CFA. A structural ESEM model was replicated in the second sample of 272 L2 students, supporting the predictive validity of the study.

A comparison of CFA and ESEM model fit has been reported by Caro et al. (2014) based on international large-scale assessment data from the Progress in International Reading Literacy Study (PIRLS) 2006 and the Programme for International Student Assessment (PISA). They evaluated and compared the factor structure of response data from item sets to measure cultural, economic, and social capital. The ESEM solution provided better-fit indices and higher support for discriminant validity in both PIRLS and PISA compared to CFA solutions for these tests. Jung (2019) compared CFA and ESEM with the application of TIMSS 2015 students' attitudes towards the science scale used. This research confirms that ESEM can provide a better representation of factor structure than CFA because it illustrates that ESEM provides a much more flexible solution over CFA’s more traditional method. In terms of model fit, factor loadings, factor correlations, and the interpretability of the model, the ESEM solution was identified as the optimal model for the TIMSS 2015 science attitudes items. The CFA and ESEM were used to assess the construct validity of the Academic Motivation Scale scores conducted by Guay et al. (2015). Their results showed that ESEM fit the data better, and the factor correlation pattern derived from ESEM was more in line with their theoretical framework. Joshanloo and Lamers (2016) also conducted CFA and ESEM with the aim of assessing mental well-being in an individual and concluded that ESEM provided more accurate factor structure results compared to CFA. In particular, it was found that the analysis of ESEM successfully distinguished two dimensions of well-being that had not been empirically distinguished in CFAs.

On the other hand, limited research shows that CFA solutions can fit better than ESEM. Gomez et al. (2020) examined the factor structure of the Depression Anxiety and Stress Scales-21 (DASS-21) in an adult community using CFA and ESEM. The researchers applied the first-order CFA, ESEM, bifactor CFA (BCFA), and bifactor ESEM (BESEM) models to compare model fit and factor structures of depression, anxiety, and stress factors. In total, 738 adults (males = 374, females = 364) completed the DASS-21 questionnaire with an average age of 25.29 years; a standard deviation of 7.61 years. Although all of the models fitted the data well in terms of the total number of factors and the number of groups and specific factors, one or more of the group-specific factors were poorly defined in the BCFA, ESEM, and BESEM.

In summary, overall evidence suggest that two of the most important features of ESEM, compared to conventional CFA/SEM, are the substantial improvement in model fit and the substantially smaller correlations between variables, both of which have been replicated in a large number of subsequent ESEM studies. For my study, this means that ESEM with its cross-loading flexibility is expected to provide a more theoretically and statistically appropriate model than CFA in terms of statistical modelling of motivational constructs.

2.3. Substantive Literature on Motivation

The section is divided into two sub-sections. An overview of motivation theories, specifically EVT, which is the theoretical framework for this study, is provided in the first part. The second part of the section presents a review of empirical literature related to motivation and achievement.

2.3.1 Theoretical approaches to motivation

A substantial body of research has investigated possible predictors of science and mathematics achievement (Bofah and Hannula, 2015; Guo et al., 2015; House, 2008; Ker, 2017; Markus and Kitayama, 1991; Meece et al., 1990; Parker et al., 2012; Shen and Tam, 2008; Wilkins et al., 2002). The study of motivation in education is one of the most attractive research interests in the literature. Notwithstanding considerable and extensive research on motivation for attainment, there is much to be learned, not only on motivation patterns across individuals but also across countries and what patterns of motivation are relevant to students' achievement across countries (Michaelides et al., 2019). Ryan (1998, p.114) stated the importance of motivation as, "If there is a cornerstone in the science of human behaviour, it must be the field of motivation. Motivational theories ask a fundamental question, namely: what motivates a person? They are concerned with the prime force at work in human nature and human culture”. The scope of motivation, in this case, covers the reason, purpose, intentions, beliefs, feelings, and attitudes about what people do and why they do things they do (Mercier and Sperber, 2017). Various theoretical frameworks have been introduced to explain the relationship between motivation to learn and achievement. Self-determination theory (Deci and Ryan, 1985), expectancy-value theory (Wigfield and Eccles, 2000), and self-concept (Marsh, 2007) are the theoretical approach of my research since the context of motivation measures in TIMSS are relevant to these theories (Michaelides et al., 2019).

The section starts with a general introduction to the motivational and related theories. Then, the theories used in the study are detailed under separate headings. Finally, the conceptual framework of the research and the adaption of theory to the current study are explained.

2.3.1.1. Self-determination theory

In the field of motivation and personality, the self-determination theory (SDT) is a broad theoretical framework concerned with people's inner growth tendencies and innate psychological needs. It helps clarify motivations behind individuals' choices on their own without external influences and intervention. This theory focuses on the degree of self-motivation and self-determination of an individual’s behaviour (Ryan and Deci, 2000;2017). Although it is quite complex in detail (see Ryan and Deci, 2002; 2017), motivation is generally defined as either extrinsic or intrinsic (Ryan and Deci, 2000).

As highlighted in the work of Deci and Ryan (1985, p.32), intrinsic motivation is an “energiser of behaviour”, with those learners encompassing motivation to learn mathematics recognising it as both enjoyable and interesting. Despite the view that all people have intrinsic motivation to learn, environmental considerations, including school and home, are seen to influence this motivation. Thus, the overall inclination to learn mathematics is whether it is considered enjoyable or not (Ryan and Deci, 2009). With this noted, other works have suggested that the extent to which a learner shows engagement in learning, and subsequently achieves performance improvement in mathematics, is affected by intrinsic motivation (d'Ailly, 2003; Tavani and Losh, 2003). On the other hand, extrinsic motivation is explained as goals or reasons for reaching an external reward (e.g. money, praise from others, good grades) or avoiding negative consequences (e.g. embarrassment for having a poor test result or not being allowed to visit a friend until the homework is completed) (Michaelides et al., 2019). Studies in the literature clearly show that intrinsic motivation is more closely related to success than extrinsic motivation (Becker et al., 2010; Vansteenkiste et al., 2008). In fact, some research suggests that external rewards reduce student intrinsic motivation (Michaelides et al., 2019). Nevertheless, few students are motivated to learn all subjects intrinsically, so teachers and parents may need to employ extrinsic rewards to motivate students. Because SDT claims that successful students internalise their extrinsic motivation to improve their performance when interacting with an environment that promotes a sense of relatedness, competence, and autonomy, such as a home or school environment (Deci and Moller, 2005; Ryan and Deci, 2000).

Attitudes and motivation have been measured through the student questionnaires in TIMSS 2019 (Hooper et al., 2013). Although there is no clear statement that SDT-guided motivation measures in terms of the operationalisation of the items, there was a reference to SDT when defining motivational structures in the recent assessment frameworks (Hooper et al., 2013; Hooper et al., 2017).

2.3.1.2. Self-concept

Self-concept can be viewed as relating to an individual's sense of self, shaped through interpretations of and interactions with one's environment and others (Shavelson et al., 1976) or otherwise referred to as "the totality of the individual's thoughts and feelings having reference to himself as an object" (Rosenberg, 1979, p.7). In this regard, self-concept acts in such a way as to create and influence actions and goals through positive or negative self-assessments, as held by a person about themselves, through their own attitudes, beliefs, and thoughts (Hattie, 1992). In this regard, one's own view of their strengths in a particular subject area is referred to as academic self-concept (Byrne and Shavelson, 1987).

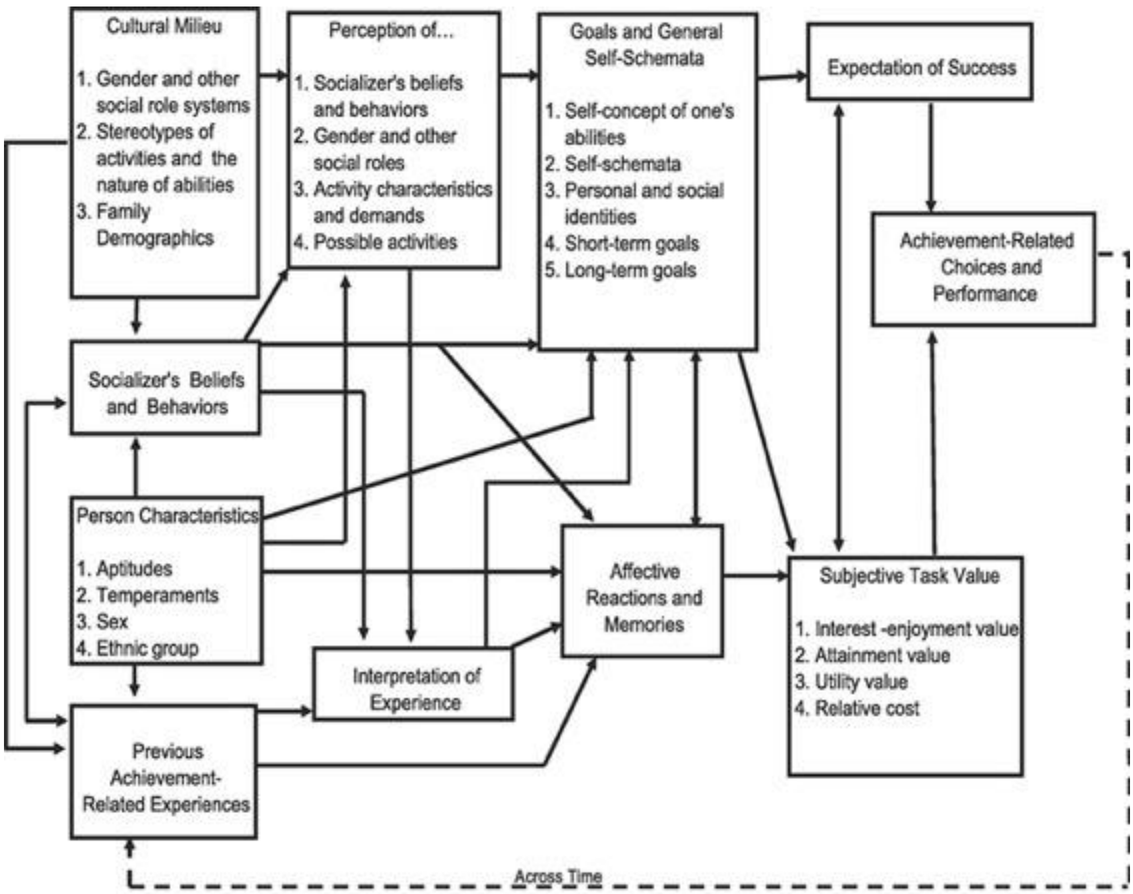
Self-concept that is positive in nature is recognised as a driving force of academic performance and attainment (Valentine et al., 2004) as well as academic effort, academic motivation, anxiety, confidence, and persistence in education, (Guay et al., 2010), academic interests (Marsh, 2007; Trautwein et al., 2006), advanced coursework selection, educational, and career aspiration, learning intentions (Eccles and Wigfield, 2020), and academic emotions, such as test anxiety, and enjoyment (Goetz et al., 2008). Furthermore, it has been stated by Guay et al. (2004) that learners' ASCs aid in predicting the educational attainment to be demonstrated by a student over a ten-year period. Moreover, the self-concept of an individual, notably in an academic domain, has been recognised as a stronger predictor of academic attainment than objective individual attainment, student background, or socioeconomic status (Guay et al., 2004; Marsh, 2007; Parker et al., 2012). In addition, the OECD (2003) stated that academic self-concept may well be linked to the economic and successful long-term well-being of a learner and should therefore be considered in line with academic attainment as a fundamental educational outcome. Self-concept is incorporated into other motivational models, such as the EVT model. Students' self-concept is measured in TIMSS by items related to their level of confidence in mathematics and science (Michaelides et al., 2019). Similar to motivational constructs (intrinsic and utility value), there is no clear information on the theoretical basis of self-concept constructs in TIMSS technical reports, but reference is made to the work of Marsh and Craven (2006) in the self-concept section.

2.3.1.3. Expectancy-value theory

When it comes to exploring the way in which motivational constructs affect and shape the choices, performance, and persistence of learners, the expectancy-value model (EVT) is recognised as being among one of the most valuable theoretical approaches. Initially, Atkinson (1957) developed the model, with Battle (1966) and Feather (1988) providing further contributions. Recently, the EVT has been developed by Eccles et al. (1983), Wigfield and Eccles (2002) and Wigfield (1994), and is a widely applied model of the EVT in the literature (Bofah and Hannula, 2015; Guo, 2016; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012).

Figure 2.2 shows Eccles et al.'s EVT model of achievement-related choices (Eccles and Wigfield, 2020). Taking the model from right to left, it appears that expectations and values have the most direct influence on achievement-related choices and performance. Meanwhile, goals and self-schemas, such as self-concepts of one's ability, affect one's expectations and values. Likewise, these goals and self-schema constructs are influenced by an individual's perception of the beliefs and behaviours of their socialisers (e.g. parents, teachers, peers) and what they have personally experienced in the past in terms of achievement-related experiences. Furthermore, individual perceptions of previous achievement-related experiences can be influenced by various cultural and social factors, such as cultural norms, gender roles, and SES of a family, as well as the individual's aptitudes, talents, personalities, and temperaments. A particular aspect of modern EVT is that it proposes causal links between achievement-related choices and previous achievements (Eccles et al., 1983). In this manner, modern EVT makes use of a developmental and integrative approach to describe the way individual and contextual influences shape students' expectations and values over time in relation to their academic choices and performance.

Figure 2. 2 Eccles' expectancy-value model of achievement choices



Source: "From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation" Eccles and Wigfield (2020).

The EVT consists of two basic components: expectancy and value. The expectancy component focuses on the question, "Can I do this task?" (that is, students' perceptions of their abilities), while a task value focuses on the question, "Do I want to do this task and why?" (Eccles and Wigfield, 2002). The expectation component in the model expresses the beliefs and judgements of the individual about their ability to perform and accomplish a task. It has clearly overlapping structures from other theoretical approaches, such as self-schemes, self-concept, or self-efficacy. The value component in the model refers to the various reasons individuals have for whether or not to participate in a task and the power of these values. Moreover, Eccles and her colleagues completed work on the value and expectancy constructs in relation to their concepts and further stated that the expectancies of learners in regard to success, ability beliefs, and subjective task values are all critical factors underpinning a learner's motivation to learn (Eccles and Wigfield, 1995; Wigfield, 1994; Wigfield and Eccles, 2000). Both of these components are considered as significant predictors of achievement behaviour (Wigfield and Eccles 1992).

Based on the original work of Atkinson (1957), and Bandura (1977), expectancy for success (ESs) is defined by Wigfield and Eccles as individuals' beliefs about how well they will do on an upcoming task, which is similar Bandura's self-efficacy concept (Wigfield and Eccles, 2000). However, Eccles et al. (1983) state a theoretical distinction between ES and individuals' self-beliefs about their current ability (e.g. ASC, Marsh (2007); self-efficacy, Bandura (1997)), because task-specific ES is considered to depend on both more general ASCs and perceptions of the difficulty of the specific tasks (Eccles and Wigfield 2020). In particular, the term "ability self-concept" is used to describe the perceptions of individuals as competent in a specific domain (Harter, 1999; Marsh, 2007), whereas the term "expectations of success" refer to one's belief in one's ability to accomplish a specific future task. It is theoretically possible to distinguish between these two types of expectations; however, since the two components show strong correlational relationships in empirical studies, self-concept and expectation beliefs are typically accepted as a single factor or as terms used interchangeably (Eccles and Wigfield, 2002). Thus, due to the issue of multicollinearity, these structures have been empirically treated as a single structure labelled as self-concept. Therefore, it is not recommended to put them as separate constructs in regression-based statistical analysis (Eccles and Wigfield, 2020). It is thus not surprising that a number of studies on EVT rely on ASC as a measurement of expectations of success (Guo, 2016; Guo et al., 2015; Marsh et al., 2013a; Musu-Gillette et al., 2015; Nagengast et al., 2011; Shen and Tam, 2008). Hence, in this study, ASC is used to measure expectancy and is refered with the phrase "expectations of success".

The value components in the EVT model are separated into four components: attainment value, intrinsic value, utility value, and cost. According to Eccles and Wigfield (2002), attainment value is defined as the personal importance of being successful at a task. Intrinsic value refers to the personal interest in a task/field or the enjoyment gained from participating in a task. If the intrinsic value is high, the reward is positive psychological results (Meyer et al., 2019). Thus, intrinsic value can be seen as intrinsic motivation in SDT due to the conceptual similarity. Utility value means the perceived (future) individual benefit of engagement and achievement in a particular task or field. Unlike intrinsic value, the utility value is recognised as being the driving factor underpinning mathematics learning owing to the fact that learners consider it to be valuable for them and their future careers (Eccles and Wigfield, 2002). Utility value address to a similar concept with the extrinsic motivational component in SDT in terms of emphasising the importance of extrinsic performance rewards. According to Eccles et al. (1983), cost refers to the perceived negative consequences of participating in a task, for example, performance anxiety leading to emotional stress. Cost also includes the perceived amount of effort required to achieve a task.

It is well documented that task value and expectancy are domain specific from pre-adolescence to early adulthood (Wigfield et al., 2009). Recent research in mathematics has found that empirical differentiations between four components of value (Conley, 2012; Gaspard et al., 2015; Luttrell et al., 2010), and these studies have shown that the value components exhibit similar correlations between them. Specifically, cost has a negative correlation with intrinsic value, utility value, and attainment value, while intrinsic value, utility value, and attainment value are positively correlated with one another.

Typically, intrinsic values are highly correlated with attainment values. Moreover, as compared to other components of value, intrinsic value tends to have a stronger correlation with expectancy. In other words, people are likely to develop competence in tasks they find enjoyable if they feel competent in them. Furthermore, both Harter (1978) competence motivation theory and Ryan and Deci's (2000) self-determination model have demonstrated that competence beliefs and intrinsic value are strongly related (Wigfield and Eccles, 2002). A number of studies have proven that the expectancy of success is a stronger predictor of academic achievement than the value beliefs whereas values are more closely related with choice behaviours, such as taking courses, engaging in academic activities, and making an effort towards achieving educational and career goals (e.g., Denissen et al., 2007; Guo et al., 2015; Wang and Degol, 2013).

Since TIMSS motivation questionnaires do not include attainment value and cost components, this thesis uses the “students like learning mathematics” and “student value mathematics” scales as a measure of MIV and MUV, which is consistent with typical applications of EVT in research that rarely includes more than two or three components (e.g., Guo et al., 2015; Musu-Gillette et al., 2015; Simpkins et al., 2012; Wang and Eccles, 2012).

After reviewing relevant theoretical literature, the following section presents empirical research on motivation and achievement.

2.3.2. Empirical research on motivation

Motivation is generally recognised as referring to students’ energy levels and inclination to learn to achieve their potential and work effectively both in school and in their behaviour. It is seen to play a key role in students’ interest in and enjoyment of school and study. Furthermore, motivation also provides a key foundation for achievement (Martin, 2001; Martin et al., 2001a;2001b; Meece et al., 1990). Students’ involvement in learning, alongside motivational constructs, is recognised as having a key effect on their achievement. Various research has demonstrated that greater engagement and motivation amongst learners leads to greater attainment overall, with both of these aspects also influencing the choices students make in regard to courses and careers. The mathematical attainment of students has been associated with learners’ activities, as well as their overall confidence and values in relation to mathematics (Mullis et al., 2012). Findings from a number of different works in TIMSS have shown that the self-confidence of a student in relation to learning mathematics is most profoundly linked with accomplishment in the subject (Guo, 2016; Kupari and Nissinen, 2013; Marsh et al., 2013a; Meyer et al., 2019). In addition, the self-concept of competence was further recognised as a key factor in terms of interpreting mathematics attainment in both high- and low-achieving regions (Wang et al., 2012). Importantly, a meta-analysis study by Schiefele et al. (1992) found a positive association between interest and achievement and career choices. The effects of academic time, attitude, and motivation on attainment in science and mathematics were examined in the work of Singh et al. (2002), with the findings illustrating the strong effects of engagement, motivation, and positive attitude on performance in both science and mathematics.

The following subsections provide a detailed literature review of the relationship between motivational constructs (MSC, MIV, and MUV), mathematics achievement, and educational aspiration. A literature review of the relationship between demographic indicators (gender and HER), motivation, mathematics achievement, and educational aspirations is also presented.

2.4.2.1. The relationship between self-concept and mathematics achievement

Shavelson et al. (1976) defined self-concept as the perception of oneself formed by the experiences associated with one’s environment. In numerous educational settings, a positive self-concept is therefore viewed as a desirable or crucial objective (Bandura, 1997; Deci and Moller, 2005; Eccles et al., 1983; Fadda et al., 2020; Marsh, 2007). Therefore, it is unsurprising that the relationship between self-concept and success has always been of interest to researchers (Bofah and Hannula, 2015; Brown and Putwain, 2022; Guay et al., 2010; Guo, 2016; Ker, 2017; Marsh, 2007; Nagengast et al., 2011; Trautwein et al., 2006; Trautwein et al., 2012; Yoshino, 2012). Due to the scope of the study, this section is limited to a review of the literature on the relationship between self-concept and mathematics achievement.

Brown and Putwain (2022) recently conducted a study of 396 upper secondary school students in England, examining the relationship between motivational factors and achievement based on average A level scores. The result of the study showed that self-concept is the highest factor in achievement among motivational factors. Moreover, self-concept not only has a direct effect on achievement but also increases the effect of other motivational factors. A study carried out in Japanese and US contexts was performed by Yoshino (2012) through the use of TIMSS (2007) data, which focused on the link between self-concept and achievements in mathematics, with the exploration taking into account the possible part played by culture in drawing comparisons between MSC in consideration to learners' parents' education levels and the number of books in a learner's home. The study results emphasised a positive link between the MSC of learners and their mathematics-related attainment across both countries' samples. The work of Choi et al. (2012) analysed the TIMSS (2007) report's background questionnaire to explore the psychological settings amongst high achievers in eighth-grade mathematics. Three factors, namely MIV, MUV, and MSC, as taken from the background questionnaires incorporated in the TIMSS, were utilised for the purpose of data comparison using a *t*-test, *z*-test, and chi-square tests for the purpose of analysis. A total of ten countries made up the context of the study, with comparisons carried out and the countries positioned in Groups A and B; the former of which comprised learners who had attained an advanced level. Group A represents an Eastern cohort of countries, while Group B is Western countries. Group A encompassed five different countries, namely Chinese Taipei, the Republic of Korea, Singapore, Hong Kong SAR, and Japan, while Group B consisted of Hungary, England, Russian Federation, Lithuania, and the United States. The study sought to provide answers to two key research questions: Were the high-achieving students of the mathematics assessment seen to possess three psychological conditions when compared with their peers, and how do high-achieving learners illustrate their trends across the three psychological conditions when compared between Groups A and B, and also Korea versus the other countries included in the sample. The study established that those learners recognised as high achievers were seen to encompass significantly more of the three psychological conditions (MIV, MUV, and MSC) in comparison to their peers. When reviewing the comparison between Korea and Group A, Korea and Group B, and Korea and the other nine countries included, no significant proportional differences were established for the MIV index; however, a significant difference at the country- and group-level was established for the other two conditions. It is interesting to note that the findings pertaining to the index MUV showed that the high achievers in Group B were far better positioned to view mathematics and its various lessons with greater importance in their present and future lives than their peers. In the review of the index MUV, another irregularity was identified with the high-achieving learners in Korea seen to demonstrate lower levels of confidence in their ability to learn and apply mathematics than their peers in Group A and all other learners in Group B. Ker (2017) examined the relationship between motivation and mathematics achievement of students in Chinese Taipei, Singapore, and the USA in a study conducted with TIMSS 2011 motivation data. According to the study's results, self-concept is the factor with the most significant effect on mathematics achievement among motivational constructs. In addition, although the self-concept values of the students in the USA were higher than those in Chinese Taipei and Singapore, their mathematics achievement was lower. Research in the literature reveals that students in Asian countries have lower motivational beliefs but higher mathematics achievement than students in Western countries. Possible reasons for this will be detailed in the discussion chapter.

A study conducted by Bofah and Hannula (2015) examined the relationship between motivational constructs and mathematics achievement utilising the TIMSS 2011, involving 38,806 students from five African countries (Ghana, Morocco, South Africa, Tunisia, and Botswana) participating in the study. The first step in their research was to investigate the psychometric properties of the mathematics motivational constructs across all five educational systems (factor structure, reliability, method effect, and measurement invariance). The TIMSS 2011 motivational construct was largely invariant across cultures, and it was empirically supported as multidimensional. The CFA also showed that negatively worded items need to be controlled in the measurement model. The results indicate that negatively worded items are treated differently in many cultures. Negatively worded items adversely affected factor structures and reliability (i.e. mathematics self-concept and mathemtaics intrinsic value). A second aspect of the study was to evaluate the relationship between the motivation constructs, the student's achievements, and the background variables. Their study found significant relationships between self-belief and maths performance, which are contrary to other work in the literature; for example, nations with high mathematics achievement seem to have students with more negative mathematics self-belief. Specifically, students' MSC tended to be less related to mathematics achievement in some countries than the value of mathematics. This situation is described in the literature as "paradoxical" (Shen and Tam 2008) or "perplexing" (Marsh et al. 2013). Despite this, consistent with cultural stereotypes, boys rated mathematics attainment higher than girls. Shen and Tam (2008) also examined the relationship between eighth-grade students' mathematics achievement and self-concept with TIMSS 1995, 1999, and 2003 data. The results of the study show that there is a strong positive relationship between achievement and self-concept within individual countries. However, a negative relationship emerges when self-concept data are aggregated at the country level. This discrepancy is probably due to differences in the perception of "self" across cultures.

The literature review showed that self-concept has a significant impact on mathematics achievement for students individually. However, cultural differences in countries show differences in the value and effects of self-concept. Therefore, this study examines the impact of self-concept on mathematics achievement in Japan, England, and Türkiye by considering cultural differences.

2.3.2.2. The relationship between task values and mathematics achievement

Researchers have found mixed results for intrinsic and extrinsic motivation. In their study about the mathematics achievement of Indian immigrants in Canada, Areepattamannil et al. (2011) showed that extrinsic motivation was negatively associated with high mathematics achievement for Indian immigrant adolescents in Canada. On the other hand, extrinsic motivation was found to be positively but insignificantly related to mathematics among Indian adolescents in India. A longitudinal study by Murayama et al. (2013) examined 3,530 German students from fifth to 10th grade. Their study found that academic achievement and extrinsic motivation are mutually related.

In a recent study, Liu and Hou (2018) analysed the National Educational Longitudinal Study 1988 (NELS88) data set, which included three waves of data, containing 1988 data for eighth graders, 1990 data for 10th graders, and 1992 data for 12th-graders. Results of the analyses indicated that extrinsic motivation for testing was a significant predictor of mathematics achievement. Furthermore, the results of the analyses indicated that mathematics performance was statistically significant in predicting the future extrinsic motivation of students. A longitudinal study conducted by Murayama et al. (2013) examined 2,530 students (Grades 5 to 10) in Germany over a period of five years. They applied latent growth curve modelling to analyse the data. As a result of the study, it was found that extrinsic motivation predicted the level of mathematics achievement in the short term but not growth in the long term; nevertheless, intrinsic motivation is a more powerful predictor for the long term.

In a study conducted by Zhu and Leung (2011), eighth-grade students from nine countries were analysed from the TIMSS 2003 data set: five East Asian countries (Hong Kong, Japan, South Korea, Singapore, and Chinese Taipei) and four Western countries (Australia, the United Kingdom, the Netherlands, and the United States). Items representing intrinsic and extrinsic motivation were defined as students' pleasure- and productivity-oriented motivations, respectively, corresponding to Deci and Ryan's (1985) concept of self-determination theory. The study showed that students from all nine jurisdictions had a significantly higher level of external than internal motivation. There was a medium difference in magnitude between the two types of motivation in East Asian systems, contrary to Western countries which had a significant difference. Put in another way, students valued mathematics more for its practical application than its enjoyment as a learning experience. In all nine educational systems, intrinsic motivation was found to affect students' achievement positively. The other noteworthy point of the study is that extrinsic motivation and mathematics achievement were positively correlated across all nine countries. The regression analyses revealed that extrinsic motivation was positively and significantly associated with mathematics achievement in Hong Kong, Japan, Korea, and Chinese Taipei. Singapore was the only East Asian country where extrinsic motivation was negatively related to mathematics achievement, but the relationship was not significant. The opposite result was the case in all four Western countries: Australia, the Netherlands, and the United States showed a meaningful negative relationship between extrinsic motivation and mathematics achievement, but the results for the United Kingdom were insignificant. The combination of both types of motivation in East Asian educational systems appears to result in higher overall motivation levels, and both types of motivation seem to work synergistically. In contrast, when intrinsic and extrinsic values are combined, extrinsic motivation becomes detrimental in the Western context or disappears.

Akben-Selcuk (2017) examined the Türkiye PISA 2012 data in a Turkish educational context. In this study, the researcher found that intrinsically motivated students tend to perform better than those with less intrinsic motivation. Additionally, her findings showed that although extrinsic motivation was positively correlated with mathematics achievement, the relationship was not significant. Kaplan (2018) examined the relationship between motivation and mathematics achievement with TIMSS 2015 Türkiye data. The results showed that intrinsic and extrinsic motivation predicts mathematics achievement positively when SES is controlled. In conclusion, the degree and direction of this relationship may vary depending on the educational and social contexts. Various types of motivation have been associated with different mathematics achievement levels, but there is no universal agreement on this point.

The literature supports this study by emphasising the importance of intrinsic and extrinsic value. It is evident, however, that cultural factors have a significant impact on intrinsic and extrinsic motivation, similar to self-concept.

2.3.2.3. The interaction between self-concept and task values

As discussed earlier, EVT was inspired by the early cognitive models of the 1940s and 1950s, which replaced earlier behaviourist models (Atkinson, 1957). One of the basic assumptions of classical EVT (Atkinson, 1957) is the multiplicative effect of expectancy and task value (i.e., expectancy-value interaction). Specifically, the multiplicative relationship between expectancy and value means that the individual value of a particular domain has an impact on the relationship between expectancy and outcomes and vice versa.

Typically, expectancy and value interactions have been described as “compensatory” or “synergistic” in terms of their effects on the outcome (Guo et al., 2016). There is a considerable difference in the nature of the interactions between the two categorisations, which can have important implications for motivation researchers from both a theoretical and substantive standpoint. Accordingly, compensatory relationships suggest that individuals will be motivated to engage in an academic task so long as they have high expectations or attach a high value to it. Basically, a high level of expectation can compensate for a low level of value, and the reverse is also true. On the other hand, a synergistic relationship suggests that high levels of task engagement are a result of both expectations of success and task value. For instance, the outcome of a task is likely to be low if a student does not expect to succeed, even when the value is high. Similarly, a low value is also likely to result in a lower outcome, even when combined with a high expectation. Nevertheless, it is important to note that although expectancy and value interaction often occurs in conjunction with substantial first-order effects (“main effects”) of expectancy and value, this complicates the interpretation of the respective effects of these two variables. It may be for this reason that previous studies have been unable to fully develop the nature of interactions pertaining to previous EVT predictions (Guo et al., 2016).

Feather (1982) reviewed papers based on original EVT models and found that expectancy and value predict a task synergistically. A significant portion of the early EVT research took place in experimental settings (e.g., Atkinson, 1957; Atkinson and Feather, 1966). Several studies have been conducted which have been conducted using a random assignment to conditions to manipulate self-concept and task value to experimentally "zero" - see Feather (1982) for a more detailed discussion.

In modern EVT, Eccles (2009, p.84) stressed that “the motivational power of ability self-concepts to influence task choice is, at least partially, determined by the value individuals attach to engaging in the domain”. The relationship between expectancy and value is often assumed to be purely additive in nature and it is often implicitly assumed that this relationship will be able to provide a unique and independent prediction of achievement-related outcomes, given that the relationship between expectancy and value tends to be additive. The possible reason for that is, in modern EVT research, non-experimental studies have become more popular over the years, which has led to greater methodological problems. The lack of interactive terms in modern research prompted Nagengast et al. (2011) to ask the question: “Who took the ‘x’ out of expectancy-value theory?” It is one of the primary objectives of this thesis to reintroduce the expectancy by task value interaction back into the EVT model by using modern statistical approaches.

In modern EVT research, the expectancy-by-value interaction has been omitted for several reasons. A first explanation may be that experimental designs which emphasised within-person (intraindividual) differences have shifted to real-world settings that emphasise between-person (interindividual) differences (Nagengast et al., 2011; Trautwein et al., 2012). Experimental research used operational definitions of expectancy and value, as well as direct manipulations of task difficulty, as key variables in their study. It is more likely that there will be a larger difference between the experimental factors when the manipulation is stronger. Nevertheless, the modern approach to EVT (Eccles, 2009; Eccles et al., 1983) considers the relationships between expectancy and value as a real-world phenomenon by linking them up to achievement-related outcomes within the context of a typical school environment, which places the relationship between expectancy and value in a practical context. As a result, experimentally manipulated differences in different tasks were replaced with naturally occurring differences in the various components of value (Busemeyer and Jones, 1983; Nagengast et al., 2011; Trautwein et al., 2012). Thus, the interaction effect of expectancy and value was examined on the basis of the differences between them that naturally occur among individuals (Trautwein et al., 2012).

Interaction effects have typically been detected as small to moderate in observational studies based on surveys or questionnaires. (Brown and Putwain, 2022; Guo et al., 2015; Meyer et al., 2020; Nagengast et al., 2011; Trautwein et al., 2012). It is possible to manipulate expectancy and value in experimental studies to more extreme levels to amplify the interaction effects (Guo et al., 2016; Nagengast et al., 2011; Trautwein et al., 2012). However, it is typically more difficult to detect interaction effects in non-experimental, empirical settings due to the lack of cases with extreme conditions (e.g. extremely high expectancy combined with extremely low task value) (Guo et al., 2016). However, in cases where expectancy and value are highly correlated, empirical studies that attempt to assess their interaction effects may be inadequate (Aiken and West, 1991).

Second, a lack of advanced statistical tools for estimating expectancy-by-value interactions may cause sparse empirical research on this topic. Even though interaction effects are detectable using manifest variables in multiple regression analysis, the effects of interaction are likely to be underestimated (Carroll et al., 2006; Marsh et al., 2013b). This is due to the fact that the predictors are measured with error, and the measurement error is multiplied when generating the product terms which makes product terms less reliable and more difficult to detect (MacCallum and Mar, 1995; Marsh et al., 2013b). For this reason, it is important to use large sample sizes and reliable predictors to prevent Type 2 errors (i.e. the detection of statistically significant interactions) (Trautwein et al., 2012).

A popular technique for controlling measurement error in non-experimental designs is structural equation modelling (SEM): a technique that uses multiple indicators to assess latent variables in order to control for measurement error. SEM provides a powerful way to detect interaction effects in non-experimental studies. Even though latent interaction models have been the subject of increasing attention since the 1980s (Jöreskog et al., 1996; Kenny and Judd, 1984; Ping Jr, 1995), they have only recently become accessible to those who wish to apply them. Among the approaches used in the field of structural equations (Klein and Moosbrugger, 2000) are latent moderated structural (LMS) equation models and unconstrained product indicator models (Marsh et al., 2004).

A number of recent empirical studies have supported latent expectancy-by-value interactions based on the latent moderated structural equation approach (LMS) (Klein and Moosbrugger, 2000) and the unconstrained product indicator approach (Marsh et al., 2004). For example, Trautwein et al. (2012) investigated the latent interactions between the four value components and the ability of ASC to promote academic achievement in a study based on German secondary school data. The results of this study demonstrate that the effects of four multiplicative terms (e.g. ASC-by value) on English and mathematic achievement are statistically significant. Nagengast et al. (2011) conducted similar research in 57 countries using the PISA 2006 data, and found a synergistic relationship between science ASC and intrinsic value and extracurricular activities and career aspirations in science. Furthermore, Nagengast et al. (2013) demonstrated that ASC and value (combination of utility value and cost) were synergistic in predicting within-person homework effort using a within-person perspective.

Guo et al. (2015) utilised EVT to examine the relationships between mathematics motivation (ASC and task values) and student background variables with Hong Kong TIMSS 1999, 2003, and 2007 data. They used latent variable models that included latent interactions to investigate the multiplicative effect of self-concept and value, both of which are fundamental components of traditional EVT. They also examined the effect of motivation and gendered tendencies as mediators. The findings indicate that: (a) self-concept is more significant in predicting educational outcomes for students with lower utility values; (b) girls and boys have comparable levels of maths self-concept and values, but girls tend to have higher maths achievement and educational aspirations, (c) family socioeconomic status plays a more significant role in determining boys' educational aspirations, and (d) the interaction of self-concept and utility value has a statistically significant effect on maths achievement and educational aspirations.

Meyer et al. (2019) aimed to replicate and extend Trautwein et al.'s (2012) study in the German federal state of Schleswig-Holstein by using a large, comprehensive sample of students (N = 3,367) who were studying at an upper secondary school in Schleswig-Holstein. They compared and tested the effect of the predictive value of expectancy-value interactions on grades, final exams, and standardised test scores as measures of achievement in English and mathematics through latent interaction modelling. According to their findings, there are measurement and domain-specific differences when predicting academic achievement using expectancy-value beliefs and interactions. It has been shown that both English and mathematics final examination results are predicted by interaction terms. Furthermore, there was a significant interaction effect for English grades, but not for mathematics.

In the literature reviewed above, expectancy-value interaction is highlighted as an important consideration. The number of studies on this subject has increased in recent years. Despite this, sufficient attention has not been paid to this issue. It is therefore the purpose of this study to provide evidence of the EVT interaction effect in Japan, Türkiye, and England, as well as determine whether this effect can be generalised across the different education systems.

2.3.2.3. Background factors

Modern EVT (Eccles, 2009; Eccles et al., 1983) assumes that individuals' performance and choices are influenced by a variety of social and cultural factors (e.g. gender and SES).

Gender Expectancy-value theory proposes that motivational beliefs play an essential role in explaining gender differences in academic choices (Eccles, 2009). There are, however, some inconsistent findings regarding gender differences in mathematics values. For example, several studies have demonstrated that boys are more likely to have higher values in mathematics (Marsh et al., 2005; Steinmayr and Spinath, 2008; Watt, 2004), whereas some have indicated no gender differences between boys and girls (Jacobs et al., 2002; Meece, J.L. et al., 1990; Wigfield et al., 1997). In a review of EVT-based research, Wigfield et al. (2016) provide a summary demonstrating how EVT can explain gender inequalities and achievement results in general. It has been found in multiple studies that males usually exhibit higher levels of maths self-concepts, attitudes, and effects than their female counterparts (Guo et al., 2015; Marsh et al., 2013a; Pampaka et al., 2011; Wigfield and Eccles, 2002). Recently, however, cross-national meta-analyses have shown that there exist gender similarities in maths achievement at a number of levels (Else-Quest et al., 2010). In addition, some research findings indicate that female students have higher educational aspirations than their male counterparts, particularly in secondary school (Schoon and Polek, 2011), whereas other studies indicate no significant differences between gender (Guo et al., 2015).

In England, recent data showed that, on average, female students achieved higher average grades at A Level than male students, despite a higher number of male students (DfE, 2019). The data are consistent with the general literature that girls perform better than boys in educational attainment (Kessels et al., 2014). According to a large meta-analysis based on an analysis of hundreds of students across elementary, middle school, high school, and university levels, girls in all subject domains had significantly higher achievement on teacher-assessed work (Voyer and Voyer, 2014). The results of large-scale assessments such as TIMSS, PISA, or PIRLS have demonstrated that girls perform better in literacy subjects than boys and underperform in mathematics (OECD, 2013). Furthermore, Kessels and Hannover

(2008) found that girls are less confident and have less interest in these fields (Eccles, 2011). It is, therefore, evident that research findings regarding the relationship between gender and achievement are inconsistent or at least vary across contexts.

It has been proposed that, according to EVT, socialisation processes can lead to shaping the conception of expectancy and value beliefs as a consequence of gender norms and roles (Eccles, 2009). The result of this is that boys acquire more favourable beliefs in domains associated with male types, such as mathematics, while girls develop more favourable beliefs in domains associated with female types, such as English (Gaspard et al., 2015). There are, as mentioned above, no consistent results to be found in the literature about gender differences in mathematics values. In a study by Jacob et al. (2002), no evidence of gender differences in mathematics values among a US sample of students from Grades 1–12 was found; however, Steinmayr and Spinath (2008) reported higher maths values for males among a German sample of students from Grades 11 and 12. On the other hand, there is evidence that gender differences exist in maths value, depending on the component of value examined (Gaspard et al., 2015). It has been indicated in various studies that although female students are aware of the importance of achieving high grades in mathematics, they have a negative perception of it as an attractive subject (Gaspard et al., 2015). The intrinsic value of mathematics has been found to be higher for males than females in German and Australian secondary schools (Frenzel et al., 2007; Watt et al., 2012). A study by Watt (2004) found that there are no gender differences in utility value from Grades 7–11, whereas Steinmayr and Spinath (2008) found male students had an advantage in 11th grade. A study examining attainment and utility value in Grades 9 and 10 found no difference between males and females in Australia, Canada, and the United States (Watt et al., 2012). The results of a study conducted by Gaspard et al.

(2015) among ninth-grade students in 25 German schools, showed that, despite the similarities between boys and girls in terms of overall value beliefs, there are gender-related differences in the mean level of various facets of mathematics value. As predicted by gender stereotypes associated with mathematics value, these differences tended to favour males. As a result, girls' intrinsic value was found to be lower compared to their male counterparts. In addition, girls tended to perceive maths as being less important in terms of their personal lives and less useful for their future in terms of their professional futures than boys. There was only one difference that favoured girls: the utility value, which is consistent with girls perceiving school and good grades as more valuable in general (Kessels et al., 2014).

According to EVT (Eccles, 2009), gender also has an effect on achievement-related behaviours through its relationship with motivational beliefs. In other words, gender differences in motivational beliefs mediate gender differences in achievement-related behaviours (Guo et al., 2015; Nagy et al., 2008; Nagy et al., 2006). A number of studies have examined motivational factors as mediating factors (Brown and Putwain, 2022; Parker et al., 2012); however, only a limited number of studies (Guo et al., 2015) have examined both self-concept and multiple task values and their interaction effects and evaluated the direct, indirect, and total effects of gender and HER on educational outcomes.

Guo et al. (2015) examined the relationship between gender, mathematics achievement, and educational aspirations with data from Hong Kong TIMSS 1999, 2003, and 2007. The results show that male students were found to have a higher mathematics self-concept and intrinsic value, but not utility value, when examining the direct effect of gender on motivational beliefs. Gender's direct impact on achievement, however, was partially counterbalanced by its indirect effect. This means that, based on the indirect path from gender to achievement via motivational factors, boys are more likely to have a high level of maths self-concept, leading to higher maths achievement levels. In contrast, girls have a higher level of mathematics achievement when their self-concepts and intrinsic values are similar to boys. In terms of the total effect, they concluded that there was no gender difference in mathematics achievement when taking all results together. However, when the mediating effect of motivational constructs on the relationship between gender and educational aspiration is considered, since male students have a higher motivation than female students, the difference in favour of girls in the relationship between gender and educational aspiration is reduced. Overall, in total effect, girls' educational aspirations were, to a small extent, favoured compared to their male counterparts.

Brown and Putwain (2022) investigated gender association with achievement through the mediation of expectancy of success, task value, and their interaction. A total of 396 participants were included in the study enrolled in their final year of upper secondary education in England. The participants completed self-report measures of their expectancy, task values, gender, and socioeconomic status. In addition, they were associated with the grades obtained in the exit examinations (A Levels). The results show that gender was not directly associated with achievement. However, gender was indirectly related to student grades through expectations, task values, and the interaction between expectations and task values.

Socioeconomic status (SES). According to the EVT framework (Eccles et al., 1983), parents have social-emotional influences on children's motivational beliefs, which in turn affect children's educational performance and aspirations. As there is a correlation between the beliefs and behaviours of parents and their SES, there are more likely to be successful outcomes for children from families with higher SES (Eccles, 2009). Numerous studies have demonstrated that the SES of a student's family plays a significant role in determining their academic success (Sirin, 2005; White, 1982). In a meta-analysis of 499 quantitative studies, Hattie (2008) concluded that SES had the most significant effect on predicting academic achievement and explained the most variance. The strong correlation between SES and achievement points to the educational inequality that has been investigated in the context of international large-scale studies such as TIMSS and PISA (OECD, 2014). Globally, comprehensive education reforms in countries have failed to mitigate the adverse effects of SES on educational outcomes, resulting in a lack of progress in educational equity (Marks, 2013).

The vast majority of research on family SES has only examined its direct and positive effects on children's academic achievement (for more detail, see Şirin, 2005), motivation (Eccles, 2007), and educational aspirations (Halle et al., 1997). In recent research, motivational beliefs have been examined as possible mediators of the relationship between SES and academic achievement and educational aspirations. (Brown and Putwain, 2022; Grolnick et al., 2009; Guo et al., 2015). For example, Guo et al. (2015), using TIMSS 1999, 2003, and 2007 Hong Kong data, found that MSC and MUV mediated the relationship between SES and mathematics achievement and educational aspirations.

A study by Kriegbaum and Spinath (2016) resolved that there was a small to moderate correlation between parents' SES and their children's mathematical achievement, that is, higher SES leads to better performance. The mediated effects of motivational factors on the relationship between parents' SES and the performance of children in standardised maths tests was also investigated. The relationship between SES and achievement was mediated by the expectancy of success beliefs (self-concept in mathematics).

Brown and Putwain (2022) investigated SES association with achievement through the mediation of expectancy of success, task value, and their interaction. The result shows that parental education was a factor directly associated with achievement. On the other hand, SES was indirectly related to student grades through expectations, task values, and the interaction between expectations and task values. Additionally, male students, those from more educated parents, and those with wealthy resources performed better in their exams, because higher expectations and higher task values also reinforce these associations. The study concluded that students' expectations of success and task values played a part in explaining the relationship between gender and SES differences in achievement.

In light of the evidence presented above and the established associations between higher SES and achievement, current research hypothesises that students with higher SES will have higher expectations and task values and this, in turn, leads to higher academic achievement.

2.4. Research Questions and Conceptual Framework of the Study

This section provides research questions and the conceptual framework of the study. The purpose of this section is to discuss how theory has been employed herein. Expectation-value theory constitutes the main theoretical framework underpinning the study. As indicated in the previous Section 2.3.1.3, titled "Expectancy-Value Theory", has a complex nature that considers achievement-related choices dependent on different factors such as cultural milieu, previous achievements, the expectation of success, task values, etc. In TIMSS background questionnaires, student motivation/attitude is measured by three different constructs. The scales “students like learning mathematics”, “students value mathematics”, and “students confident in mathematics” were developed under the general title of students’ attitudes towards mathematics. It should be noted that the theoretical framework of TIMSS 2019 motivational measurements is not clearly stated in the report (Hooper et al., 2017). Although the theoretical basis on which these scales are based is not explicitly stated, in the report prepared by Hooper et al. (2017), self-determination theory (Deci and Ryan, 1985) was referred to for the “students like learning mathematics”, and “students value mathematics” scales, and Marsh and Craven (2006) ASC theory for the “students confident in mathematics” scale. In this thesis, “students like learning mathematics”, “students value mathematics”, and “students confident in mathematics”, used in TIMSS 2019, measures are applied as an indicator of EVT factors: MIV, MUV, and MSC, respectively. This theory divides task value theoretically into four facets: intrinsic value, utility value, attainment value, and cost. The TIMSS study, however, arguably only contains structures that represent intrinsic and utility values (Hooper et al., 2017). Due to data limitations, statistical estimation concerns, and the broad concept of EVT, it is not possible to apply all the factors in a single study. Therefore, this thesis uses only expectation of success/self-concept, intrinsic value, and utility value, which is consistent with typical applications of EVT in research (e.g. Guo et al., 2015; Musu-Gillette et al., 2015, Simpkins et al., 2012; Wang and Eccles, 2013).

Furthermore, as demographic indicators, gender and student HER variables were selected based on the EVT framework. Specifically, the SES of students was measured via the “home educational resources (HER)” scale in TIMSS. The scale consists of three items, namely, “the number of books”, “other study supports in their homes”, and “the highest level of their parents’ education” (Hooper et al., 2017). Therefore, the variable “HER” is considered as an indicator of SES in this study.

First, the study compares the psychometric properties, such as factor structure, negative items effect, and measurement invariance, of TIMSS 2019 motivation variables based on traditional CFA and ESEM. This analysis aims to evaluate the validity and reliability of motivation items and the measurement invariance across Japan, Türkiye, and England to determine the best model for the data. Even though TIMSS data is commonly applied to investigate the relationship between student motivation and attainment (Choi et al., 2012; Guo et al., 2015; Ker, 2017; Yoshino, 2012), there are surprisingly few studies that emphasise the psychometric issues and invariance of the latent factors (Bofah and Hannula, 2015; Marsh et al., 2013a). Further, to the best of my knowledge, comparative CFA and ESEM model fits of different countries using TIMSS 2019 mathematics motivation data have not yet been conducted in the literature. In this regard, one of the objectives of this research is to contribute to the existing literature to fill the gap in the evaluation of psychometric properties of TIMSS 2019 motivation measures and CFA vs ESEM discussion.

The results of the psychometric analysis are presented in Chapter 4, titled “ Evaluation of Psychometric Properties of TIMSS 2019 Motivation Measures and Their Relations With Demographic Variables and Educational Outcomes”. For this part of the study the following research questions are investigated.

(1) What are the psychometric properties of TIMSS 2019 motivation measures?

a) Do TIMSS 2019 motivation variables provide reliable measurement properties for Japan, Türkiye, and England?

b) Do negatively worded items affect model fit?

c) Is the ESEM analysis superior to CFA in the measurement model of TIMSS 2019 motivation latent constructs?

d) Do TIMSS 2019 motivation measures support the a priori factor structure for the Japan, Türkiye, and England samples?

e) Do TIMSS 2019 motivation variables have measurement invariance among the countries?

(2) What are the correlational relationships between motivation factors, gender, HER, and educational outcomes (mathematics achievement and educational aspirations)?

a) Does the relationship between motivation factors, gender, HER, and educational outcomes vary across the countries?

b) Is there a difference between the latent means of motivation factors across the countries?

After determining the best model fit and ensuring measurement invariance, I move to more substantive questions by examining the relationships of motivational factors with mathematics achievement and educational aspiration. As discussed above, the interaction effect of self-concept and task value was emphasised in the original EVT (Atkinson, 1957), but in modern EVT (Eccles et al., 1983), it has been omitted (Nagengast et al., 2011). Therefore, in this section, in addition to the direct effects of EVT factors on mathematics achievement and educational aspiration, the interaction effects of self-concept and task value are also analysed. Figure 2.3 presents the conceptual framework for this analysis. This conceptual framework was created within the framework of modern EVT (Eccles et al., 1983; Eccles, 2009), and the interaction of expectancy and task value discussed in the literature is added based on the literature (Nagengast et al., 2011; Trautwein et al., 2012). In other words, while the arrows from MIV, MSC, and MUV to outcome variables (MAT ACH and EDU ASP) seen in Figure 2.3 are based on modern EVT, the arrows from MSC to the path from MIV and MUV to outcome variables represent the interaction that has been criticised in the literature.

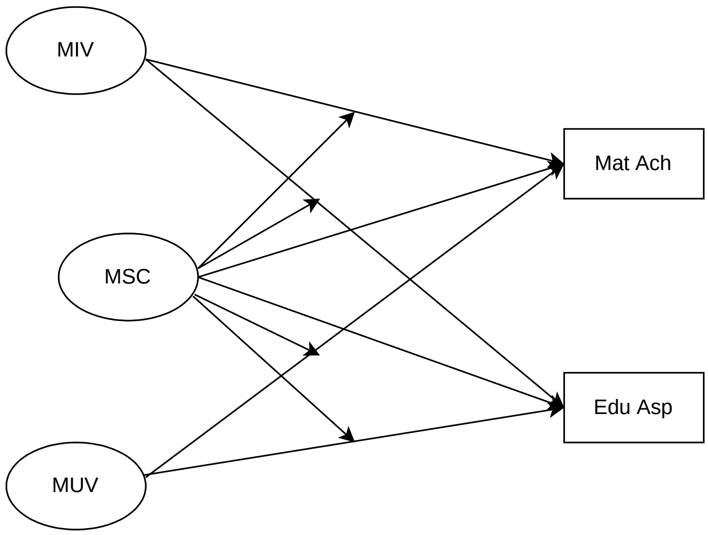
The results of this substantive analysis section are presented in Chapter 5, titled "The Relationship Between Expectancy-values and Their Interactions With Educational Outcomes". The following research questions are analysed for this chapter.

(3) What is the relationship between student motivation and educational outcomes (mathematics achievement and educational aspirations)?

   a) How well do motivation factors predict educational outcomes?

   b) Is there an interaction effect between expectancy and value beliefs on educational outcomes?

Figure 2. 3 The hypothesised conceptual framework of EVT factors and their interaction in relation to mathematics achievement and educational aspirations



Note. MIV = mathematics intrinsic value MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement; EDU ASP = educational aspirations; the arrow from MSC to the path of MIV and MUV to the outcome variables represents the moderation/interaction.

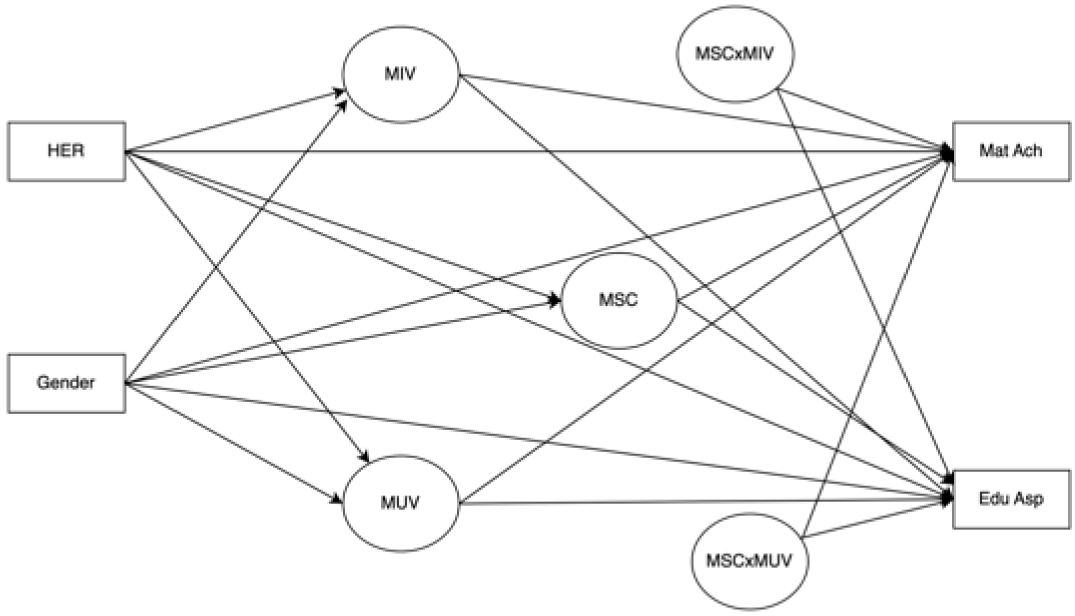
Although EVT theory emphasises the direct and indirect effects of SES and gender, few studies have investigated the complex relationship between SES, gender, MSC, MIV, MUV, mathematics achievement, and educational aspiration through mediation (Guo et al., 2015; Brown and Putwain, 2022; Parker et al., 2012). The last part of this thesis aims to contribute to this gap by investigating how SES and gender are related to mathematics achievement and educational aspiration mediated by task values and MSC. In the EVT (section 2.3.1.3), Figure 2.1 shows the theoretical approaches of the EVT model on the indirect effects of personal characteristics and cultural milestones, family character, and gender on achievement-related choices and performance. Based on this theoretical framework, the study's conceptual framework is adapted (as shown in figure 2.4). Finally, the results of this section are presented in Chapter 6, titled " The Mediation Effect of Expectancy-Value factors on Educational Outcome".

As a result, the following research questions are investigated:

(4) What are the direct and indirect effects of gender and HER on mathematics achievement and educational aspirations through the mediation of EVT factors in Japan, Türkiye, and England?

(5) What is the relationship between HER mathematics achievement and educational aspirations? Do EVT factors mediate this relationship?

Figure 2. 4 The hypothetical conceptual framework of the mediating effect of EVT in the relationship between background factors and mathematics achievement and educational aspirations



Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement; EDU ASP = educational aspirations; MSCxMIV = interaction variable of MSC and MIV; MSCxMUV = interaction variable of MSC and MUV.

2.5. Summary

This chapter consists of two main sections. The first contains literature related to the SEM and factor analyses, which are the main analysis methods of this study. In this section, first, the traditional CFA method for modelling latent data and the more recent ESEM models, which have recently been introduced to the literature but have limited studies, are compared. Accordingly, literature review on psychometric evaluations such as measurement invariance, factor structure, validity, and reliability are presented. In the second section, theories and empirical studies related to motivation are presented. The second section starts with theoretical approaches to motivation, mainly EVT, and provides a theoretical framework for this study. The reason for the use of EVT, briefly, is because it does not examine the concept of motivation in terms of task values (intrinsic and extrinsic motivation) only. However, at the same time, the individual's expectation of success, a broad concept encompassing the elements of self-concept and perceived task difficulty, relates task value and achievement to related choices and performance. Therefore, since the interaction effect of self-concept and task value is one of this thesis's potential key contributions, EVT is employed in the current study. The effects of EVT components on educational outcomes are presented based on empirical studies in the literature. Finally, the literature review chapter ends with the "research questions and conceptual framework of the study" section, where I explain the study's research questions and how these questions are conceptualised based on the EVT and related literature.

Chapter 3: Data and Methods

3.1. Introduction

There is a recognised importance concerning the completion of cross-cultural works on mathematics-related research regarding the way in which different factors could impact current thinking and further contribute to determining more universal and valuable theories (Chen, 2005; Marsh and Köller, 2004). These works should centre on improving understanding not only in regards the differences between countries' mathematics attainment factors but also in relation to the measured constructs' psychometric properties of latent variables. In line with this, the study consists of two main parts: the first (Chapter 4) focuses on evaluating TIMSS 2019 motivation measures methodologically, the second (Chapters 5 and Chapter 6) aims to examine the relationship between motivational factors and educational outcomes and background variables. Marsh and Hau (2007) have described this approach as "substantive-methodological synergies", joint ventures where new methodological developments are applied to important issues or new methodological approaches are designed to provide better answers to core research questions. On this basis, the aim of this study can be summarised with a single statement as follows: In essence, this study aims to explain the relationship between motivational factors by applying the modern and appropriate methodological approaches (methodological part detailed in Chapter 4) and to analyse the motivation factors (expectancy and value beliefs), which are among the significant factors that impact student mathematics attainment within the EVT framework (substantive part detailed in Chapters 5 and 6).

This chapter is divided into two main sections: data and methods. The data section begins with a description of the data source, sampling and plausible values and then details the variables and missing variables in the study. In the methods section, I first explain the methodological approaches of the study. In the conclusion section, I summarise the research questions and the statistical analysis method used for each research question.

3.2. Data

3.2.1. Trends in International Mathematics and Science Study (TIMSS) 2019

The TIMSS is an international assessment project arranged by the IEA (International Association for the Evaluation of Educational Achievement). The projects have been conducted on a four-yearly basis since 1995 across grade levels four and eight, with the most recent being carried out in 2019. It further delivers comparative viewpoints on patterns concerning attainment in the context of different educational systems and frameworks, instructional practices, and organisational approaches. This is achieved through TIMSS gathering a wealth of different background information (Hooper et al., 2017).

The data source in this instance is the TIMSS (2019), with the data recognised as valuable when performing an analysis of the research questions for several reasons. Primarily, the data were gathered internationally from a large sample population, but as a secondary consideration, a two-stage stratified probability sampling was devised in TIMSS bearing in mind the choice of learners and schools, while the random sampling process was carefully carried out (LaRoche et al., 2020). Third, through the TIMSS 2019, subjects were questioned about their attitude towards learning and their backgrounds.

In this work, the subjects, as detailed below in Table 3.1, are in Grade 8 in Japanese, Turkish, and English schools. The subjects completed the student background questionnaire.

Table 3. 1 Number of participants for each country

Countries	Students	Classroom	Schools
Japan	4,446	142	142
Türkiye	4,077	181	181
England	3,365	161	136

3.2.2. Sampling weights and clustering

The TIMSS data utilized in this thesis demonstrate a hierarchical structure, with students nested within schools and classes. To effectively implement a true multilevel Structural Equation Model (SEM), it is essential to disentangle the variance existing within school and between school/class levels, thereby enabling the simultaneous modeling of relationships at different levels (Guo, 2016). The primary objective of this thesis is to examine the relationships among latent variables at the individual level, excluding any consideration of school/class-level relationships. Consequently, opting for a single-level SEM appears to be appropriate and satisfactory (Stapleton, 2013). However, disregarding the sampling design effects present in the clustered sample data may lead to biased estimates of standard errors (Stapleton, 2013).

To address this issue, the thesis adopts complex design modelling, effectively resolving the problem. This approach yields the same parameter estimates of path coefficients as single-level modelling while appropriately accounting for the nesting of students within the school/class level, thus ensuring corrected standard errors for the parameter estimates. The implementation of this method is achieved in Mplus 8 through the utilization of the function TYPE = COMPLEX (Muthén & Muthén, 1998-2017). By employing the complex function in Mplus, corrected standard errors for the estimates are provided, along with a scaled chi-square statistic that is robust to the non-independence of clustered observations within the same cluster (Stapleton, 2013).

A two-stage complex survey design used in TIMSS, which brings different probabilities for different level units (LaRoche et al., 2020). In the first sampling stage, schools were selected based on probabilities proportional to size from a list of all schools in the population. In addition, the school list can be stratified based on significant demographic variables. The second sampling stage consisted of one or more entire classrooms, randomly selected in the school (LaRoche et al., 2020). Unlike the sample selection of PISA, in which the number of 15-year-old students was randomly selected from among the entire school population, TIMSS chose one or more classes from selected schools. All students in the selected classes were subject to assessment.

In addition to two-stage sampling, student weighting methods are also applied to accurately reflect the participating countries' population in TIMSS. Individuals are weighted differently by their sampling weights, reducing bias caused by stratification, non-responsiveness, or subsamples' disproportionality (Ker, 2017). National student samples in TIMSS are designed within a specified sampling error margin to represent the target populations accurately. After TIMSS 2019 data was collected, student characteristics' means and percentages were weighted based on each country's population to estimate population parameters (LaRoche et al., 2020). To make it more straightforward, what sample weighting is and what it is used for can be explained with the following example:

> Assume that we are interested in the study habits of a particular classroom with 18 students, 12 boys and 6 girls. If we randomly choose 6 students to participate in our study, we would expect to select 4 boys and 2 girls on average. Assume, however, that it is important to include an equal number of boys and girls in our study while accounting for the fact that girls represent a smaller proportion of students in our hypothetical class. Using this approach, we would select 1 girl for every 2 students surveyed, giving each girl a 3/6 probability of selection. Similarly, we would also survey 1 boy for every 2 students selected, implying a 3/12 probability of selection for each boy because we would choose 3 of 12 possible boys. To ensure that girls are not overrepresented in our resulting estimates, every surveyed student's response is adjusted to reflect the student's actual proportional occurrence in the population. These adjustments are the sampling weights (Rutkowski et al., 2010, p.143).

For student-level data analysis, TIMSS 2019 provides three weighting variables: total student weight (TOTWGT), student senate weight (SENWGT), and student house weight (HOUWGT). The TOTWGT is the overall student sampling weight, which is the sum of the student population size in each country; SENWGT is a linear transformation of TOTWGT to reach an equal sample size of 500 for each country; and HOUWGT is also linear transformation of TOTWGT and based on the actual number of students in the sample that is appropriate for the correct computation of standard errors and tests of statistical significance (Rutkowski et al., 2010). In this study, all analyses are based on TIMSS's HOUWGT weighting variable, which includes three components related to sampling of the school, class, and student and three related to non-participation at the school, class, and student levels. Therefore, the use of sampling weights in the analysis provides proper correction for clustering inherent in the two-stage clustering sample (Marsh et al., 2013a). For the present purpose, country variables were considered as grouping identification in multigroup analyses to control the clustered sample. Class ID was used as the clustering variable rather than school ID because, as mentioned above, the sampling unit was based on class in the TIMSS sampling design. A complex design option is applied to correct the standard error caused by the two-stage sampling design (see Muthén and Muthén, 1998-2017).

3.2.3. Plausible values

To represent a full repertoire of knowledge of students' mathematics skill and strategies, TIMSS consists of many questions for students, thus, to provide good coverage, they need to have up to 10 ½ hours of testing time (Fishbein et al., 2019). However, due to time limitations and difficulties, the assessment time for each student booklet must fit into 90 minutes for the eighth grade by clustering student items in blocks and randomly rotating the blocks of items through the 14 student test booklets. The questions included in these booklets were carefully selected from the TIMSS (Trends in International Mathematics and Science Study) question pool. This selection process took place collaboratively, incorporating the curriculum survey results submitted by the participating countries to TIMSS. As a result of this meticulous procedure, 14 booklets were developed and are implemented in the TIMSS measurement program by the respective executive institutions in the participating countries. The main advantage of using this method is that all the questions in the question pool are put to students randomly. Consequently, although not all students answered all the questions, each item was responded to through randomly designed item booklets by students (LaRoche et al., 2020). This approach or design is called a matrix sampling approach/design. Item response theory scaling method is used to produce a broad result of the achievement of the whole student population in a country from the combined answers of individual pupils to booklets (LaRoche et al., 2020).

On the other hand, using a matrix sampling design comes at the cost of an inability to make statements at an individual level. A considerable amount of uncertainty accompanies individual proficiency estimates, and simply aggregating individual scores will result in seriously biased demographic profiles (Haneuse and Bartell, 2011).

In order to overcome this issue, plausible value methodology was developed by (Mislevy, 1991). International tests such as TIMSS, PIRLS, and PISA use plausible values to assess student achievement, as measurement errors may arise in educational studies due to many factors, such as mental and physical circumstances on the assessment day, and the existence of potential conditions influencing the outcome of the assessment (Mullis et al., 2020). Specifically, "the plausible values are not test scores … but are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual" (Monseur and Adams, 2009, p.6). Plausible values are neither actual student scores nor imputed scores for individuals, but instead "imputed scores for like students with similar response patterns and background characteristics in the sampled population" to provide an accurate estimation of population. The logic of the method is to use all available data, including background data, to estimate the characteristics of population and sub-populations and to use multiply imputed scores, called plausible values to account for the uncertainty.

The TIMSS 2019 data consists of five plausible values representing an overall achievement measure in mathematics for each student. The analyses' results are calculated based on the combination of five plausible values using the MPlus (Muthén and Muthén, 1998-2017) function "*type = imputation*" following Rubin (1987) multiple imputations procedure.

3.2.4. Variables used in the study

The TIMSS collects information such as students' motivation and demographic school environment through the student background questionnaire, and measures students' mathematics and science achievements, which addresses their attitudes towards mathematics as well as their interest in science. Along with the outcome variables (mathematics achievement and educational aspirations), student attitudes towards mathematics and home resources for learning were selected from the student questionnaire as a requirement of this study hypothesis.

The TIMSS survey (TIMSS 2019) encompasses several different background and contextual variables, all of which could have an influence on the links between academic attainment and mathematics performance, including classroom and teacher factors, demographics, school conditions, and management. Notably, however, this work has selected only some of these so as to enable the work to provide a clearer emphasis. The variables used in this study were operationalised under three main categories; educational outcomes, expectancy-value constructs, and background variables. A 4-point Likert scale was used in all motivation items and scaled from 1 "disagree a lot" to 4 "agree a lot", in which higher scores represent more favourable responses.

3.2.4.1. Expectancy-value measures

Expectancy-value. The MSC scale was used to evaluate students' expectancy-value. The scale consisted of nine items in TIMSS 2019 (e.g. "I am generally good at mathematics"). Items of self-concept variables can be viewed in Table 3.2.

Task values. The scales of "students like learning mathematics" measure the intrinsic value of a student to learn mathematics. Mathematics intrinsic value is assessed with nine items (e.g. “I enjoy learning mathematics”). Utility value is measured through scales of “students value mathematics” in TIMSS 2019. Mathematics utility value is assessed with nine items (e.g. “I think learning mathematics will help me in my daily life”), each with a 4-point response format ranging from “disagree a lot” to “agree a lot”. The items of intrinsic and utility values are represented in Table 3.2.

Table 3. 2 Items of expectancy-value factors

Latent Variables	**Item Wording**	**Response scale**
Mathematics Intrinsic Value (MIV)		
MIV1	I enjoy learning mathematics	
MIVN2*	I wish I did not have to study mathematics	
MIVN3*	Mathematics is boring	
MIV4	I learn many interesting things in mathematics	
MIV5	I like mathematics	
MIV6	I like any schoolwork that involves numbers	
MIV7	I like to solve mathematics problems	
MIV8	I look forward to mathematics class	
MIV9	Mathematics is one of my favourite subjects	1= Disagree a lot 2= Disagree a little 3 = Agree a little 4= Agree a lot
Mathematics Self-Concept (MSC)		
MSC1	I usually do well in mathematics	
MSCN2*	Mathematics is more difficult for me than for many of my classmates	
MSCN3*	Mathematics is not one of my strengths	
MSC4	I learn things quickly in mathematics	
MSC5	I am good at working out difficult mathematics problems	
MSC6	My teacher tells me I am good at mathematics	
MSCN7*	Mathematics is harder for me than any other subject	
Mathematics Utility Value (MUV)		

MUV1	I think learning mathematics will help me in my daily life
MUV2	I need mathematics to learn other school subjects
MUV3	I need to do well in mathematics to get the job I want
MUV4	It is important to learn about mathematics to get ahead in the world
MUV5	Learning mathematics will give me more job opportunities when I am an adult
MUV6	My parents think that it is important that I do well in mathematics
MUV7	It is important to do well in mathematics

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2019 Copyright © 2020 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College. * Negatively coded items

3.2.4.2. Outcome variables

Mathematics achievement. In this study, the mathematics achievement scores obtained from the mathematics test of Japanese, Turkish, and English eighth-grade students who participated in TIMSS 2019 were used. Item response theory (IRT) was used to scale scores derived from students' tests to evaluate achievement and obtain accurate measures in TIMSS (Hooper et al., 2017). Since each student is only responsible for parts of the assessment item pool, the TIMSS scale approach obtains proficiency scores using the plausible value/multiple imputation methods (Fishbein et al., 2021).

Educational Aspirations. Students' long-term educational aspirations were measured with the single item as "How far in school do you expect to go?" in TIMSS 2019. The response scale consists of six answers ranging from "Finish lower secondary education" to "Finish postgraduate degree: master's or doctor". Education levels by country are shown in Table 3.3. In this study the measure of educational aspirations is treated as continuous scale variable.

3.2.4.3. Background/demographic variables

Gender. Gender was self-reported and coded 1 for girls and 2 for boys, so higher coefficients indicate higher scores for boys (see table 3.3).

Home educational resources (HER). This was assessed with a scale including three items which are the highest educational level of father and mother, the number of books at home, and the number of home study supports (see table 3.3). The scale is created with Item response modelling.

Table 3. 3 Details of outcome and background variables

Variables	Description	Details of Categories
MAT ACH	Imputation of five plausible mathematics achievement values	-
EDU ASP	How far in education do you expect to go?	**Japan** 1 = Lower secondary school 2 = Upper secondary school 3= Advanced course of upper secondary school 4 = Junior college, college of technology, or specialised training college (post-secondary course) 5 = University or college 6 = Graduate school **Türkiye**

		1 = Lower secondary education 2 = Upper secondary education 3 = post-secondary vocational courses 4 = Short-cycle tertiary education 5 = Bachelor’s level 6 = Master’s or doctor **England** 1 = Lower secondary school up to the age of 14 2 = GCSE, AS, A level, or equivalent qualifications, e.g. NVQ at Level 3 or GNVQ 3 = Higher education access course 4 = Higher education qualification below degree level e.g. NVQ Level 4 or 5, diploma, nursing qualification, or higher level in HNC, HND, or BTEC 5 = University degree (e.g. BA, BSc, BEd) 6 = Master’s degree, doctorate or higher degree (e.g. MPhil, PhD)
Gender	Student gender	1= Female 2= Male
Home educational resources (HER)	Students were scored according to their reports regarding the availability of	-

		number of books and study supports in the home and their parents' education level

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2019 Copyright © 2020 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College.

3.2.5. Missing data

A total of 4,444, 4,077, and 3,365 students were involved in TIMSS 2019 for Japan, Türkiye, and England, respectively. However, observations of any item to which the participant did not respond (2 from Japan, 29 from Türkiye, and 158 from England) were excluded from the analysis. Therefore, a total of 4,442 observations from Japan, 4,048 from Türkiye, and 3,207 from England were included in the analysis. The multiple imputation method was used for the remaining missing data. The number of missing data by country and variables is shown in Table 3.4. As seen in the table, there are much fewer missing data in the case of Japan compared to Türkiye and especially England.

Although the percentage of missing values is not high, many variables contain a missing value. With increased awareness of the limitations of listwise deletion, mean substitution, or pairwise deletion, recently methods such as full information maximum likelihood (FIML) and multiple imputations have become more popular (Marsh et al., 2013a). In this work, multiple imputations are applied to handle missing data. Five imputed data sets were created for each plausible value, so 25 data sets were obtained. One of the five data sets of each plausible value was kept for the analysis. The multiple imputations strategy allowed me to use plausible values properly and deal with missing data in the same data set simultaneously. All variables used in this study were added in imputation models, and models weighted using the HOUWGT weighting variable. The results reported in this study are based on an appropriate aggregation of results from multiple data sets to obtain the appropriate parameter estimates, standard errors, and goodness-of-fit statistics according to the Rubin (1987) multiple imputation automatic procedure by MPlus 8.6 (Muthén and Muthén, 1998-2017).

Table 3. 4 Number of available and missing data

	Japan		**Türkiye**		**England**	
	N	***Miss***	***N***	***Miss***	***N***	***Miss***
MIV1	4,442	2	3,974	74	3,172	35
MIVN2	4,437	7	3,955	93	3,155	52
MIVN3	4,425	19	3,942	106	3,116	91
MIV4	4,430	14	3,949	99	3,152	55
MIV5	4,430	14	3,948	100	3,135	72
MIV6	4,440	4	3,929	119	3,163	44
MIV7	4,442	2	3,962	86	3,150	57
MIV8	4,434	10	3,966	82	3,155	52
MIV9	4,442	2	3,969	79	3,158	49
MSC1	4,437	7	4,007	41	3,153	54
MSCN2	4,435	9	3,997	51	3,129	78
MSCN3	4,429	15	3,977	71	3,096	111
MSC4	4,427	17	3,984	64	3,112	95
MSC5	4,434	10	3,992	56	3,131	76
MSC6	4,424	20	3,974	74	3,094	113
MSCN7	4,433	11	3,993	55	3,135	72
MUV1	4,437	7	4,007	41	3,142	65
MUV2	4,436	8	4,002	46	3,141	66
MUV3	4,420	24	3,993	55	3,130	77
MUV4	4,429	15	3,979	69	3,129	78
MUV5	4,422	22	3,997	51	3,125	82
MUV6	4,419	25	4,001	47	3,121	86
MUV7	4,433	11	4,005	43	3,127	80
MAT ACH 1	4,444	0	4,048	0	3,207	0
MAT ACH 2	4,444	0	4,048	0	3,207	0

MAT ACH 3	4,444	0	4,048	0	3,207	0
MAT ACH 4	4,444	0	4,048	0	3,207	0
MAT ACH 5	4,444	0	4,048	0	3,207	0
HER	4,438	6	4,044	4	3,183	24
EDU ASP	4,393	51	3,969	79	3,086	121
MALE	2,276	0	2009	0	1,715	0
FEMALE	2,168	0	2039	0	1,492	0

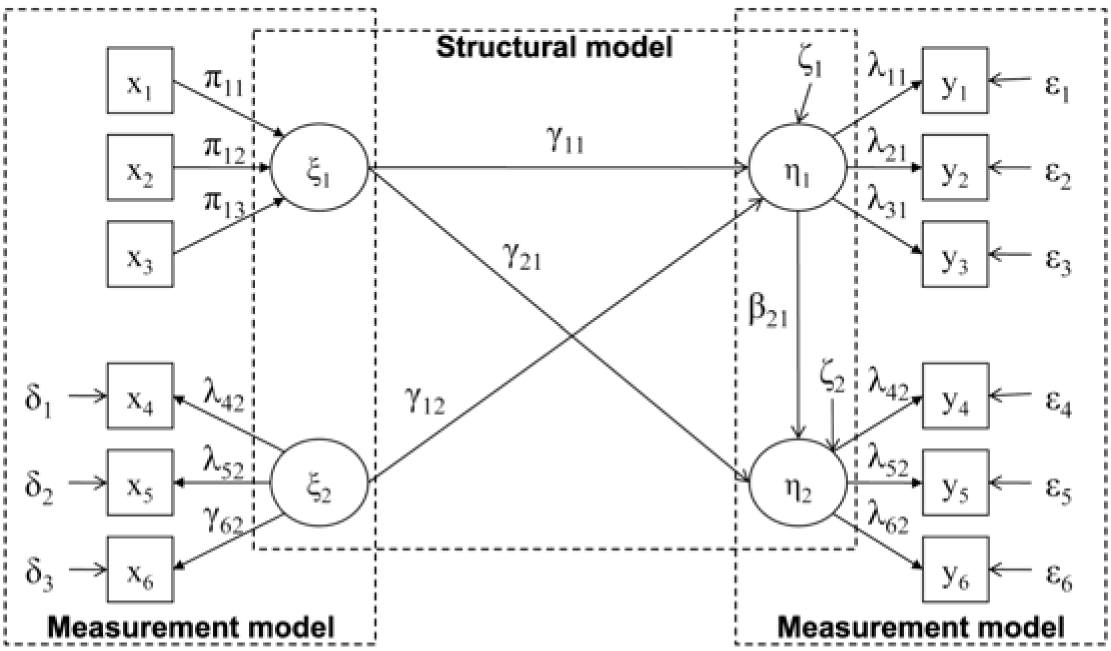
NOTE. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement score; HER = home educational resources; EDU ASP = educational aspirations.

3.3. Methods

The main analysis method of this study is SEM. Structural equation modelling is an umbrella term that includes multiple statistical models widely used in social and behavioural sciences (Bollen, Kenneth A., 1989; Hoyle, 2012). These models include, for example, multivariate regression, path analysis, factor analysis, growth curve models, and multilevel (regression, factor, and path) modelling. The SEM method is often applied to analyse models containing both measured and structural parts. The measurement model, path model, and SEM example are shown in Figure 3.1. The measurement part relates the latent variables of the model to their manifest indicators. The structural part assumes hypotheses about the relationships between a set of variables of interest, some of which may be latent. The main concern of analysis in many practical settings is the structural part as it reflects the theory that the researcher is testing. The measurement part only occurs when the model includes latent variables that must be measured using (typically multiple) manifest indicators.

Confirmatory factor analysis (CFA) is probably the most common method used in the SEM framework to examine the test structure and measurement model (Hoyle, 2012). However, due to some limitations of CFA, which I will detail in the next section, the ESEM approach was developed by Asparouhov and Muthén (2009). Since the ESEM approach is relatively new compared to CFA, its use in the literature is limited. In this study, I will first examine the psychometric properties of TIMSS 2019 motivation factors, such as factor structure, measurement invariance, model fit, validity, and reliability, based on a comparison of two robust measurement model approaches, CFA and ESEM. In the literature, some scales which were developed to measure motivation previously tested for validity and reliability with the CFA approach are also retested with ESEM (Alamer, 2021; Alamer and Marsh, 2022; Fadda et al., 2020; Furnham et al., 2013; Gomez et al., 2020; Jung, 2019; Marsh et al., 2013c; Morin et al., 2020; Morin et al., 2013; Tóth-Király et al., 2017; Xiao et al., 2019). However, there is no study, at least to my knowledge, in which TIMSS 2019 motivation measures were retested and modelled with the ESEM approach and compared with CFA. Therefore, in this respect, the present study aims to make a unique contribution to the methodology literature with an ESEM and CFA comparison, in addition to a detailed examination of the psychometric properties of TIMSS 2019.

Figure 3. 1 Example of measurement and structural model



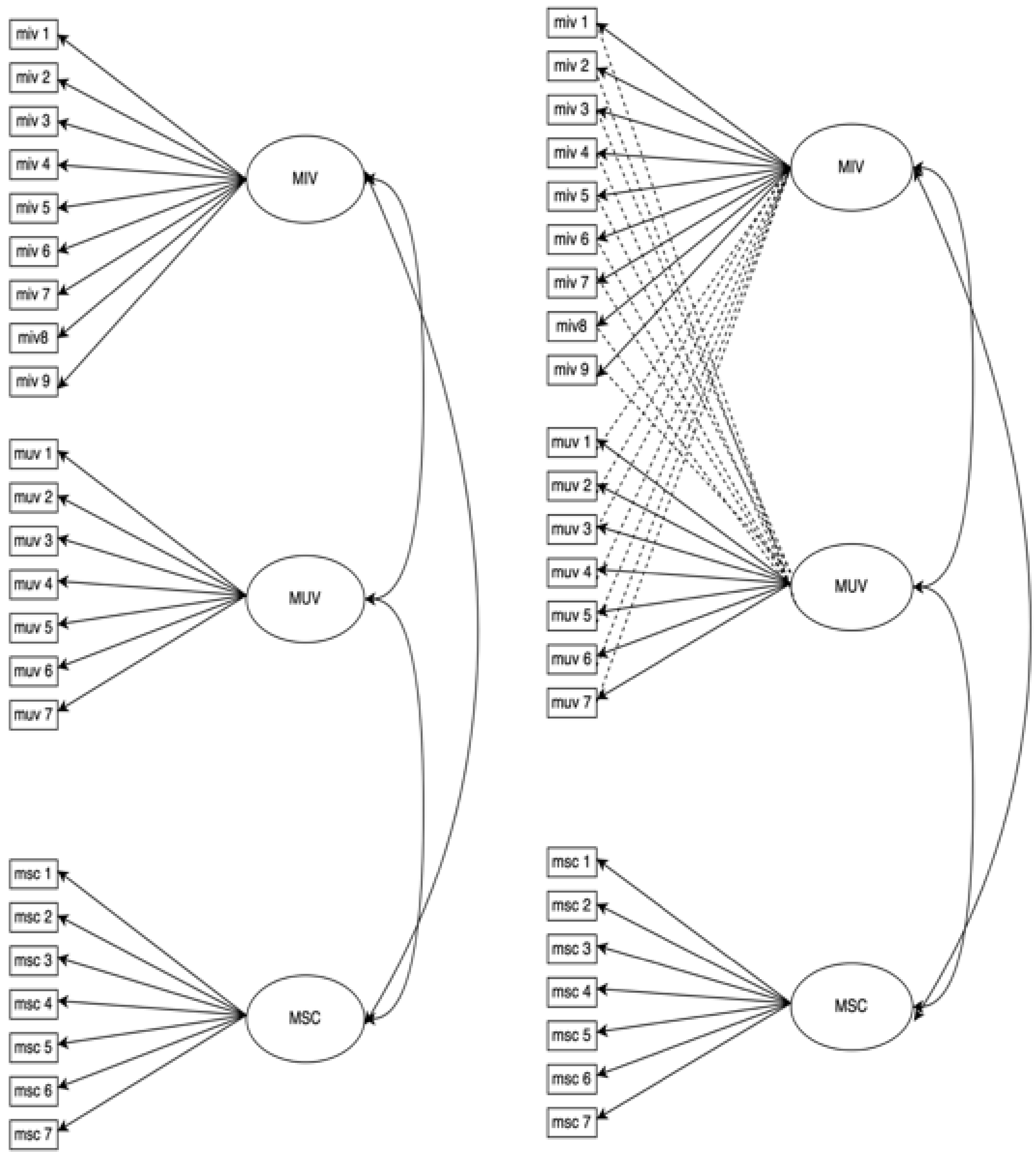
Source cited from Roldán and Sánchez-Franco (2012).

In the analysis of structural models section (section 3.4), SEM models which are developed according to the results of the measurement models, are explained. This section discusses the relationship between motivation, mathematics achievement, educational aspirations, and background factors from different perspectives. First, the SEM approach that includes product indicators is applied to discuss the first-order and interaction effects of expectancy and value (EVT) factors on educational outcomes. The last section examines the direct, indirect, and total effects of gender and HER factors on mathematics attainment and educational aspirations through EVT components with mediation analysis.

3.3.1. Measurement model: a comparison of CFA and ESEM

As mentioned in Section 2.2 of Chapter 2, in an ESEM model, items can be loaded on both main factors (solid lines) and other factors with target rotation as close to zero (dashed lines) (see figure 3.2). Factors and items that are allowed for cross-loading should be theoretically related to each other to properly explain and evaluate the association between the factors (Marsh et al., 2020). This study discusses motivational factors within the EVT framework. Therefore, we have two primary constructs: expectancy and value. While the expectancy structure was measured within the framework of self-concept, the value structure was measured with two sub-factors as intrinsic value and utility value in TIMSS 2019. Although the full ESEM model gives better-fit statistics when the expectancy and value constructs are allowed to be cross-loaded, it is not theoretically accurate since expectancy and value factors are supposedly separate constructs (Marsh et al., 2020). Thus, in this study, in the ESEM model, self-concept and value constructs do not to be need cross-loaded; following Marsh et al. (2020), only value constructs (intrinsic and utility) are allowed to cross-loaded within themselves. In this context, the CFA and ESEM models developed for this study are shown in Figure 3.2. This study uses these models to examine factor structure, measurement invariance, reliability, validity, method effect of the motivational constructs and their relationship with educational outcomes, and background factors using TIMSS 2019 Japan, Türkiye, and England data through these models. These psychometric properties are explained in detail under the following subsections.

Figure 3. 2 Graphical illustrations of measurement models used in the study.



a) CFA Model b) ESEM Model

NOTE. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value.

3.3.2. Method effects: negative item effect

Marsh and colleagues define method effects thus, “Method effects are non-trait effects associated with idiosyncratic aspects of particular items or methods of data collection” (Marsh et al. 2013a, p.112). Failure to include method effects appropriately can significantly affect the goodness of fit, biased parameter estimates, and meaningful interpretations.

Scales measuring psychological constructs are vulnerable to various responses such as involuntary approval or careless answers during the implementation of an instrument (Cronbach, 1946, cited in Michaelides, 2019, p.366). Including both positive and negative statements in the structure has been suggested to reduce these effects (Michaelides, 2019). The idea behind it is to create a cognitive mechanism where the role of negatively worded items requires participants to have more control than automatic cognitive processing (Podsakoff et al., 2003). However, the literature does not support these claims; on the contrary, it has been found that responses to negative statements produce systematic variations that are not related to the content studied, mostly in younger age groups but including all age groups (Benson and Hocevar, 1985; Marsh, 1986).

In this study, I examine the method effect caused by using a mix of positive and negative statements within the same construct used to measure motivational factors in TIMSS 2019. There are four negative items in total, two each in self-concept and intrinsic motivation (see table 3.2).

The correlated traits and correlated uniquenesses (CTCU) framework is a commonly used methodological approach to counter the method effect under SEM (Bofah and Hannula, 2015; Guo et al., 2015; Guo et al., 2016; Marsh, 1986; Marsh et al., 2013a).

The CTCU framework constructs a correlation among the error terms in a negatively worded item so that the negative item effect is more conceptualised as a methodological artefact rather than a discrete factor (Ye and Wallace, 2014). Hence, the CTCU framework does not examine the negative items as a unique factor to remove the negative item effect (Distefano and Motl, 2009). This aims to eliminate the irrelevant empirical association between variables or constructs from the study.

Therefore, the method effects associated with negatively worded items are hypothesised based on the information provided above. It is, therefore, necessary to check for method effects in data in order to obtain a model with adequate model fit and parameter estimates.

3.3.3. Measurement invariance

The measurement invariance is used in the CFA/ESEM context to test whether the measured constructs mean the same thing for groups. The measurement invariance is necessary to make a meaningful comparison between the group scores (Marsh et al., 2013a). Measurement invariance is tested in three main steps: "configural, weak factorial (also known as a metric), strong factorial (also known as scalar)" (Widaman and Reise, 1997). This study does not compare manifest scores, so the strict measurement invariance test does not apply.

Configural invariance. The measurement invariance test starts with the configural invariance model. This step tests whether the constructs used in the analysis are in the same pattern in the tested groups by freely estimating all the parameters. A lack of configural invariance may arise due to cultural differences when constructs developed for one culture are adapted to other cultures (Chen, 2008).

Metric invariance. If a good model fit is reached in the configural model, the second step is the metric invariance test. Metric invariance, also referred to as *weak measurement invariance* or *factor loading invariance*, measures whether the factor loading of items is invariant among the groups. Achieving metric invariance indicates that the constructs have the same direction for factor loadings in each group (Millsap and Olivera-Aguilar, 2012). Metric invariance failure can be caused by various reasons, such as different meanings of measured structures for other groups, translation errors, and differences in participants' responses to items due to varied cultural backgrounds. (Chen, 2008).

Scalar invariance. Scalar or strong measurement invariance requires that intercepts and factor loadings of indicators are invariant across groups. Scalar invariance is usually tested after metric invariance is established because differences in factor loadings indicate that the regressions of the measured variables on factor scores are not parallel across groups (Millsap and Olivera-Aguilar, 2012). This is because if these regressions are variable, group differences at measurement intercepts are also possible, as two regression lines with different slopes (i.e. factor loads) typically have different intercepts (Millsap and Olivera-Aguilar, 2012). It is essential to establish scalar invariance when the means of variables are compared between groups, as this implies that differences in the means of variables must be caused by factors common to across groups (Millsap and Olivera-Aguilar, 2012). Thus, it removes an important source of uncertainty in explaining group differences in means (Chen, 2008).

Strict invariance. This requires the invariance of factor loadings and intercepts and item uniqueness. Strict measurement invariance is necessary to compare manifest scale scores (or factor scores) because reliability differences for multiple groups will distort the mean differences in observed scores (Marsh et al., 2013a). However, scalar invariance support is sufficient for a valid comparison of latent means for comparisons based on latent structures. (Marsh et al., 2013a). It does not require the additional assumption of measurement error invariance. Since the comparison of manifest scores is not part of this study, the strict measurement invariance test is ignored here. In this study, three countries are considered as a grouping variable for multigroup analysis.

3.3.4. Reliability

Cronbach's alpha (α) has been widely used as a reliability indicator test statistic for many years (Novick and Lewis, 1967). However, there is evidence that imported constructs have lower reliability when applied to different cultural contexts, regardless of their high reliability in their original settings (Bofah and Hannula, 2014). Furthermore, α may be underestimated or overestimated when correlated measurement errors are present in the underlying structure (Brown, 2015; Raykov, 2012). Raykov (2012) suggests that the composite reliability (ω) measure provides more stable results in CFA/SEM models and should be used to support α estimates. This means that a composite reliability estimate is more precise than one provided by α in cases where method effects are present, such as correlation uniqueness due to negative items (Brown, 2015; Raykov, 2012). In exploratory research, values greater than 0.600 are also recommended as acceptable levels of reliability (Hair et al., 2019). According to our hypothesis, the reliability of the estimates would be lower in Japan than in Türkiye and England since the constructs are largely derived from Western research and empirical evidence from previous TIMSS studies (Bofah and Hannula, 2015; Marsh et al., 2013a; Metsämuuronen, 2012; Rutkowski and Rutkowski, 2010). In addition, I hypothesise that the method effect associated with the negatively worded item would adversely affect the reliability of the MSC and the MIV constructs (Bofah and Hannula, 2015; Marsh et al., 2013a).

3.3.5. Convergent and discriminant validity

An intriguing aspect of the discourse surrounding the theoretical framework of self-motivation centres on the methodologies proposed for validating this framework. While confirmatory factor analysis (CFA) has been commonly employed to assess the adequacy and compatibility of models, thereby confirming the theoretical soundness of psychometric measures, recent studies in the field suggest that exploratory structural equation modelling (ESEM) could offer a promising alternative to CFA approaches (Fadda et al., 2020; Kocur et al., 2022; Marsh et al., 2020; Marsh et al., 2010; Marsh et al., 2009). This is primarily attributed to the way ESEM deals with model specification and non-target loadings within the models.

When utilizing CFA, researchers are required to predefine the parameters of the proposed model concerning the associations between observed and latent constructs. Typically, this approach leads to a situation where each indicator of the latent variable is exclusively linked to only one factor, with other possible loadings being constrained to zero. In contrast, ESEM necessitates only information on the number of latent factors, allowing for free estimation of the other parameters. In ESEM, all factors have the potential to be linked to all indicators, similar to exploratory factor analysis (EFA) (Marsh et al., 2020). As a result, the ESEM procedure enables the estimation of cross-loadings, which provides supplementary insights into the theoretical validity of the scale.

The ability to observe cross-loadings and the flexibility offered by ESEM have led to the notion that ESEM might present a more suitable framework for analysing the factor structures of psychological inventories intended to measure complex constructs of individual differences, such as personality traits and motivational factors (Kocur et al., 2022).

Importantly, the ESEM approach includes target rotation, granting the researcher a priori control over the hypothesized factor structure. This involves assuming that cross-loadings, if feasible, are close to zero but not strictly equal to zero, which distinguishes it from the CFA. Consequently, ESEM can be employed in a confirmatory manner, unlike EFA, which is commonly used for exploratory purposes in understanding potential theoretical structures of a construct. In this context, ESEM serves as a valuable means of validating the theoretical framework of a given construct (Kocur et al., 2022). Considering these latest insights from existing literature, the current study evaluated the validity of the latent constructs employing the multiple-indicator-multiple-indicator cause (MIMIC) model (M11) under the ESEM approach.

The TIMSS motivation factors are further evaluated using the extended MIMIC models by relating the latent motivation factors to pupils' background variables (gender and HER) and educational aspirations, and mathematics achievement scores. This study uses a MIMIC approach where each of four correlates (gender, HER, educational aspirations, maths achievement) is related to latent motivation factors. The MIMIC approach is similar to multiple regression, but considerably more robust because it is based on latent constructs that are purged of measurement errors and controlled for method bias, rather than simply assuming the model is accurate (Marsh et al., 2013a). In addition to latent motivational factors, variables gender, HER, educational aspirations, and mathematics achievement have been included in this model. It is hypothesised that academic achievement is more strongly correlated with self-concept than task value, and educational aspirations are more strongly correlated with task value than self-concept (e.g. Eccles et al., 1983; Eccles and Wigfield, 2002; Marsh et al., 2005; Wigfield et al., 2009).

3.4. Analysis of Structural Models

In this study, SEM is used to test the theoretical model using TIMSS 2019 data. It is widely recognised that SEM is one of the most powerful multivariate analysis methods in the social sciences (Hoyle, 2012). It is a technique for multivariate statistical analysis that uses a confirmatory (hypothesis-testing) approach in order to examine the relationships between a variety of variables, both observed and unobserved, and provide precise estimates of the errors as well as the direct and indirect effects of the variables under analysis. A further benefit of SEM is that both independent and dependent variables can be considered latent constructs (Kline, 2015). A wide range of analyses can be performed with SEM, from simple to complex relationships between variables. The framework allows researchers to use empirical models to test the validity of a theory by developing and analysing complex relationships among multiple variables. A major advantage of the method is that it allows for the management of measurement error, which is one of the biggest weaknesses of most studies (Hoyle, 2012; Kline, 2015).

In this study, I compared two measurement models for TIMSS 2019 motivation measures and found that ESEM provides a better solution than CFA in terms of model fit and other psychometric properties. Therefore, the ESEM model was retained for the structural model to test whether the self-concept and task value were predictive of educational aspirations and maths performance.

First, the task value variables (MIV and MUV) were added to the structural model, followed by the self-concept variables. This was followed by the self-concept and intrinsic value interaction variables and, finally, the self-concept and utility value interaction values.

The main hypotheses of our study are as follows. It is expected that the mathematics self-concept and intrinsic value would be highly correlated based on both conceptual considerations and previous empirical findings. The second hypothesis is that self-concept and value beliefs are positive predictors of achievement and educational aspirations when entered separately into the model. Third, when both expectancy and value beliefs are included in the regression equation, I expect expectancy beliefs to be the stronger predictor of academic achievement. The fourth objective of the study is to examine whether expectancy and value beliefs predict achievement and educational aspirations additively or synergistically. It is my primary interest here to determine whether or not the interaction term indicates a synergistic/multiplicative relationship.

3.4.1. Mediation analysis

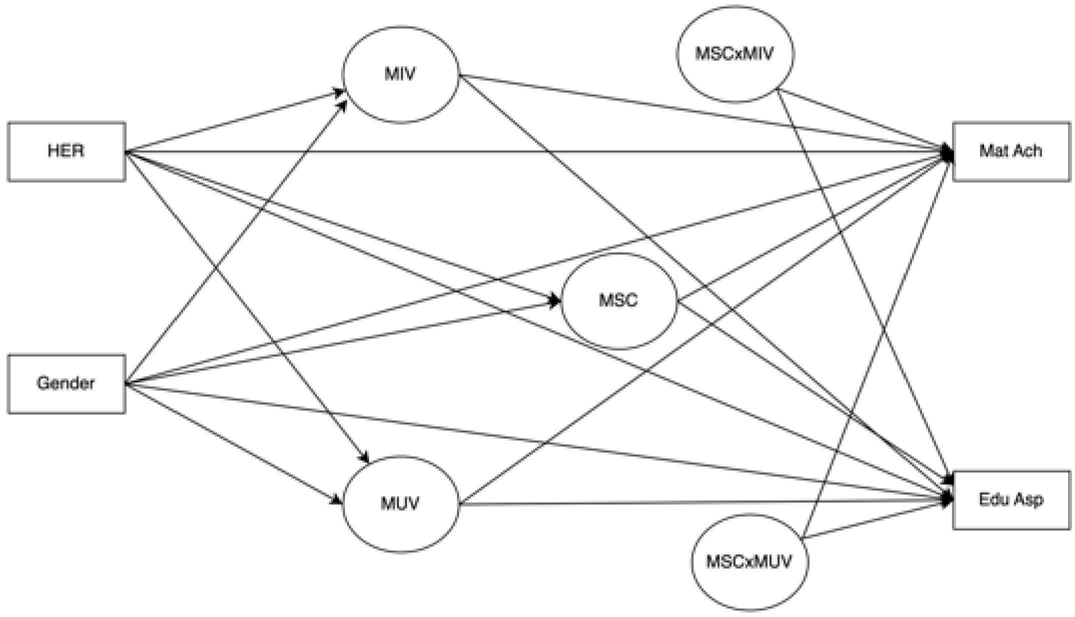
Mediation models are statistical models that attempt to explain observed relationships between independent and dependent variables by including a third hypothetical variable known as a mediator variable (MacKinnon, 2008). Mediation models propose that the independent variable influences the mediator variable, which in turn affects the dependent variable rather than being a direct causal relationship. Thus, the mediation variable provides insight into the relationship between the independent and dependent variables (MacKinnon et al., 2007; VanderWeele, 2016). Mediation analysis with cross-sectional data is a statistical technique used to investigate the underlying mechanism through which an independent variable (IV) influences a dependent variable (DV) by operating through one or more intermediate variables known as mediators. This analytical approach aims to explore the indirect effect of the IV on the DV, as well as the direct effect, thereby shedding light on the potential causal pathways involved in the observed relationships (Shrout and Bolger, 2002).

Mediation analysis with cross-sectional data can provide valuable information about the indirect relationship between variables, but it cannot determine causality. In other words, statistical mediation analyses using nonexperimental data offer suggestive indications rather than conclusive proof concerning causal relationships. The findings can propose potential causal mechanisms but cannot definitively establish causality due to the limitations of the study design and potential confounding factors (Shrout and Bolger, 2002). To establish causality, longitudinal studies or experimental designs are required.

Through a mediator variable, mediation analyses examine how one variable influences another to understand a well-known relationship. It is particularly useful when there is no obvious direct relationship between independent and dependent variables (MacKinnon, D.P. et al., 2007).

The purpose of this part of the study is to examine the mediation role of motivational factors. Motivational factors (MSC, MIV, and MUV) act as a mediating variable in this model, where gender and HER are the independent variables. This analysis aims to examine the mediating role of motivation factors in the relationship between demographic variables (gender and HER) and educational outcomes (maths achievement and educational aspiration). It is hypothesised that self-concepts and task values would mediate the relationship between SES, gender, and educational outcomes. The hypothesised structural model is presented in Figure 3.3.

Figure 3. 3 A hypothesised final structural model with mediating relationships and interaction terms.



Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement score; HER = home educational resources; EDU ASP = educational aspirations. MSCxMIV = interaction term of MSC and MIV; MSCxMUV = interaction term of MSC and MUV.

3.4.2. Estimation

All analyses in the study were performed using MPlus version 8.6 (Muthén and Muthén, 1998-2017) with IBM SPSS (version 26) for data screening. The analyses are based on the MPlus robust maximum likelihood estimator (MLR) with standard errors and fit tests, which are robust for the use of categorical variables with at least four or more response categories, especially when the non-normality of the observations is not extreme (Marsh et al., 2013a). On this basis, considering the complexity of the models used in the study, I chose to use MLR estimation which treats Likert responses as continuous variables instead of a categorical estimation procedure. This decision is based on studies in the literature (Muthén and Muthén, 1998-2017) and research conducted with the MLR estimation method (Bofah and Hannula, 2015; Guo et al., 2015; Marsh et al., 2013a), which suggests that the categorical estimation method has little or no difference to parameter point estimate, especially when extensive as in the current study. In addition, the skewness and kurtosis of the items and composite variables in the study are not excessive (Hau and Marsh, 2004); the average skewness is –0.76 (none more than 1.7), and the average kurtosis is -0.22 (kurtosis of only one item is greater than 2 in absolute value).

Another reason for using the MLR estimation method in this study is to investigate the complex relationships of variables in SEM models (moderation and mediation analysis). Because the WLS estimation method used for analyses of categorical variables in mediation and moderation analyses is not available in Mplus 8 version. Accordingly, the programme developers recommended using MLR estimation for a related question on the MPlus blog site (Muthen, 2015). In addition, studies have shown different advantages and disadvantages of using MLR and WLS in ordinal data. For example, in simulation studies, it has been observed that factor loading of ordinal data with DWLS estimator is better and less biased (Bandalos, 2014), while MLR produces a less biased standard error, more accurate intercorrelation and structural model (Li, 2016). Current research practice lacks agreement on preferred estimation methods when dealing with observed variables of varying scale types. However, although the use of MLR in a technical sense is seen as a limitation in this study, considering that the primary purpose of the study is to examine the structural relationship with mediation and moderation analyses, using MLR as an estimation method offers more advantages than WLS as recommended by Li (2021). In conclusion, upon comprehensive examination of the strengths and drawbacks associated with assuming ordinal variables as continuous variables, it becomes evident that this approach may present theoretical limitations, notably concerning non-normal distribution. Nonetheless, the implementation of the MLR estimation procedure allows for the mitigation of such non-normality concerns, rendering it more suitable and relevant for this study to treat Likert scale-type ordinal variables as continuous variables. This is particularly advantageous, considering the need for conducting mediation and moderation analyses in the context of this research.

3.4.3. Model evaluation criteria and model fit

Multiple criteria were applied for model fit evaluation. Since the chi-square test is sensitive to large sample sizes, it can be problematic as a single model evaluation criterion (Marsh et al., 1988). Thus, applied CFA and SEM research focuses on model indices that are independent of the sample size, such as root mean square error of approximation (RMSEA), the Tucker-Lewis index (TLI), and the comparative fit index (CFI) (e.g. Marsh et al., 2013a; Bofah and Hannula, 2015; Guo et al., 2015; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012). The two indexes, TLI and CFI, can take values in the range of 0 to 1, and values above 0.90 are acceptable; values over 0.95 are considered to be a better model fit (Hu and Bentler, 1999). Meanwhile, RMSEA values below 0.8 and 0.6, respectively, are considered acceptable and a good model fit (Hu and Bentler, 1999). However, in model comparisons for nested models, the increase or decrease in model fit indices is more important than any model's absolute fit level (Kline, 2015; Marsh et al., 2013a). Cheung and Rensvold (2001) suggested that there is reason to support the more parsimonious models when the decrease in model fit for these models is less than 0.01 for CFI, and increases by less than 0.015 for RMSEA. Although these guidelines are used as a reference in the present study, it should be kept in mind that these are only general guidelines (Marsh et al., 2004) and can vary across studies (Kline, 2015). Marsh et al. (2013a) recommended that applied researchers should consider different indices such as chi-square, detailed evaluations of the real parameter estimates related to the theory, a priori predictions, comparison of viable alternative models, and common sense to evaluate the goodness of fit.

3.5. Ethics

This thesis is based on the analysis of secondary quantitative data, which ensures that schools and students participating in the study remain anonymous throughout the entire process. As secondary data was used in this study, the University of Leeds light touch ethical protocol was applied, and a copy of the ethics report can be found in Appendix 1. The research conducted in this project has adhered to guidelines established by the University of Leeds (UoL, 2021).

Secondary data analysis involves using existing research data to answer different research questions from the primary study (Tripathy, 2013). Secondary data can be data collected for large-scale surveys or personal research. One of the main concerns with the secondary use of data is the possibility of harming the people involved and ensuring that their consent is respected (Tripathy, 2013). However, the analysis excludes human subjects if the data do not identify participants' personal information (UCONN, No Date). In the TIMSS 2019 data used in this study, the participant's personal information was kept confidential by TIMSS, and this information was presented to the researchers coded with numbers. Therefore, this study's participants' information is anonymous and considered an exemption for human subjects.

There are some ethical responsibilities associated with the use of secondary data by the researcher. First, it is important to ensure that the data are made publicly available in the main data source, and it remains anonymous to protect personal information's privacy and confidentiality. The reader should also be informed about the data used in the study in the necessary level of technical detail. Further, the data must be suitable for the study's purpose and hypotheses and not be manipulated to achieve the desired outcome.

3.6. Summary

This chapter consists of two main sections describing the data set used in the study and the method of the study. In the first part, TIMSS 2019 data for Japan, Türkiye, and England are explained, while in the second, the statistical approaches used to analyse these data are presented.

As stated in the introduction and literature chapter, this study aims at substantive-methodological synergy (Marsh and Hau, 2007). In this context, the first part of the analysis compares CFA and ESEM to find the most suitable factor analysis methodologically. Therefore, the method section introduces factor analysis, factor structure, measurement invariance, negative item effect, and reliability analysis. Then, it details how substantive analyses are applied, specifically model development, latent interaction, and mediation models of SEM models.

The following table (Table 3.5) provides a summary of the research questions, literature, theoretical bases, purpose, main analytical approaches, and chapter of analysis.

Table 3. 5 An overview of the study's research questions, their basis, their purpose, and methods of analysis

Theoretical/Literature Bases	**Research Questions**	**Purpose**	**Analysis Methods**	**Results**
TIMSS was designed in a Western context (Bofah and Hannula, 2015; Marsh et al., 2013a; Metsämuuronen, 2012; Rutkowski and Rutkowski, 2010). Mathematics attitudes instruments have been critiqued for their weakness in justifying validity, particularly in non-Western countries (Abu hilal, 2001; Bofah and Hannula, 2015; Marsh et al., 2013a). ESEM superiority over traditional CFA in terms of model fit, factor corrections and theoretical representations (Marsh et al., 2014).	1) What are the psychometric properties of TIMSS 2019 motivation measures?	To determine the most suitable model for the analysis of TIMSS 2019 motivation structures under SEM.The evaluation of psychometrics properties such as factor structure, measurement invariance, reliability, and method effects of TIMSS 2019 motivational constructs by comparing Japan, Türkiye, and England	Factor analysis, measurement model, measurement invariance, reliability test. Measurement model comparison based on CFA and ESEM model.	Chapter 4
A concern with TIMSS motivation data is the possible high correlation between maths intrinsic value and maths self-concept, which threatens the convergent and discriminant validity of constructs.	2) What are the correlational relationships between motivation factors, gender, HER, and educational outcomes (mathematics achievement and educational aspirations)?	To test the correlation between TIMSS 2019 motivational constructs with each other and the correlations of these constructs with educational outcomes (mathematics achievement and educational aspirations) and demographic indicators (gender and HER) to examine whether they support convergent and discriminant validity.	Multiple indicators multiple causes (MIMIC) model	Chapter 4

Task value (mathematics intrinsic and utility values) will positively predict educational aspirations, whereas mathematics self-concept should be the strongest predictor of mathematics achievement (Eccles and Wigfield, 2002; Marsh et al., 2013a; Simpkins et al., 2006). Although EVT emphasises the effects of cultural factors on academic motivation and performance, most tests with this model have been conducted in Western cultures (Wigfield et al., 2004). In addition, expectancy and value interaction is one of the main elements of original EVT (Atkinson, 1957), but is not used in modern EVT (Eccles, 1983).	3) What is the relationship between student motivational factors and educational outcomes (mathematics achievement and educational aspirations)?	To investigate the effectiveness of EVT factors for predicting mathematics achievement and educational aspirations in different education systems (Japan, Türkiye, and England). A further objective would be to examine how expectancy and task value interaction, which has disappeared in modern EVT, predict mathematics achievement and educational aspirations.	Structural equation modelling with latent interaction with unconstrained approach.	Chapter 5
In the EVT framework, gender and SES are linked to educational performance (Eccles, 2007, 2009). Since parental behaviour and attitudes are related to SES, students from high SES families tend to perform better academically (Eccles, 2009). Similarly, gender indirectly affects academic performance through its relationship with motivation (Eccles, Barber, and Jozefowicz, 1999; Brown and Putwain, 2022; Guo et al., 2015). The mediating role of motivation factors has been discussed in the literature (Brown and Putwain, 2022; Guo et al., 2015; Parker et al., 2012), but to the best of my knowledge, no study has examined how expectancy,	4) What is the relationship between gender mathematics achievement and educational aspirations? Do EVT factors mediate this relationship? 5) What is the relationship between HER mathematics achievement and educational aspirations?	As noted, few studies have investigated the complex relationships between gender, SES, MSC, task values (MIV and MUV), mathematics achievement, and educational aspirations in a single model. In this part of the thesis, I examine the relationship between SES and gender with regard to educational aspirations and mathematics achievement, as well as the interaction between MSC and task value. The main objective of this analysis is to extend previous empirical studies by exploring how these factors explain the complex	Structural equation modelling with mediation model.	Chapter 6

task value, and their interaction effects influence mathematics achievement and educational aspirations from a comparative perspective.	Do EVT factors mediate this relationship?	picture of mathematical achievement and educational aspirations.		

Chapter 4: Evaluation of Psychometric Properties of TIMSS 2019 Motivation Measures and Their Relations with Demographic Variables and Educational Outcomes

4.1. Introduction

In the first three chapters, the introduction, literature review, and data and methods chapters are presented, respectively. The results of this study are presented in three separate chapters. This chapter mainly focuses on the measurement models and psychometric features of latent variables based on the comparison of CFA and ESEM approaches while the second and third results chapters present predictive SEM models results which include mediation and moderation models that focus on explaining the relationship between motivational factors, attainment, and demographic factors. These chapters are organised taking into account the research questions and analysis methods.

This chapter develops as follows; First, the relevant research questions are restated and research hypotheses are explained. Next, I present the results of the analysis in the following section. This section is divided into four subsections: descriptive statistics, reliability of the TIMSS 2019 motivation scales, factor structure of TIMSS 2019 mathematics motivation scales, and construct validity of TIMSS 2019 mathematics scales: relations to correlates. Finally, the summary section is presented.

The result section starts by giving descriptive statistics of variables used in this study. Next, the psychometric properties of motivation measures, such as factor loadings, measurement invariance, and reliability are discussed as a result of the comparison of the CFA and ESEM models. Finally, a MIMIC model is developed by adding both educational outcome and demographic variables as covariates into the model to evaluate the convergent and discriminant validity of motivational factors and their relations with covariates based on correlation analysis. My objective in this chapter is to determine which statistical modelling approach is more accurate (CFA vs ESEM) and also to evaluate the psychometric aspects of TIMSS 2019. The results of this analysis are important to help in the design of the predictive models for the chapters to come.

4.2. Research Questions

Here the relevant research questions and associated hypothesis are presented.

(1) What are the psychometric properties of TIMSS 2019 motivation measures?

a) Do TIMSS 2019 motivation variables provide reliable measurement properties for Japan, Türkiye, and England?

H1: Based on the wider literature, I expect motivation data will meet the reliability requirement in all three countries; however, since the questionnaire was developed in Western countries, England's reliability indexes will be higher than that of Japan and Türkiye.

b) Do TIMSS 2019 motivation measures support the priori factor structure for the Japan, Türkiye, and England samples?

H2: I expect responses to motivation items will support three a priori factors (mathematics intrinsic value, mathematics self-concept, and mathematics utility value) in all countries.

c) Is the ESEM analysis superior to CFA in the measurement model of TIMSS 2019 motivation latent variables?

H3: Models based on ESEM will provide better model fit indices for motivational factors data than models based on CFA.

d) Do negatively worded items affect model fit?

H4: Negatively worded items will negatively affect the model fit indices for CFAs and ESEM models.

e) Do TIMSS 2019 motivation variables have measurement invariance among the countries?

H5: TIMSS 2019 data from Japan, Türkiye, and England will be able to provide configural, metric, and scalar invariance.

(2) What is the correlational relationship between motivation factors, gender, HER, and educational outcomes?

a) Is there a difference between the latent means of motivation factors across the counties?

H6: The latent means of motivation factor will differ significantly between countries and Japanese students are the least motivated across the three countries.

b) How does the relationship between motivation factors, gender, HER, and educational outcomes vary across the countries?

H7: Motivational factors will be correlated with gender, HER, long-term educational aspirations, and mathematics achievement scores across all countries. MSC and MIV will also have a strong correlation.

4.3. Results

4.3.1. Descriptive statistics

Descriptive statistics are presented in Table 4.1. The tabulated data presents descriptive statistics for the variables under scrutiny. It is important to acknowledge that MIV, MSC, and MUV are latent variables, and the displayed skewness, kurtosis, and mean values in the table were derived from the averaging of scale items. Additionally, detailed individual distributions of the scale items can be found in Appendix 2. This appendix further provides distribution graphs for the variables denoted as home educational resources (HER), educational aspirations (edu asp), and mathematics achievement (mat ach). In this study, the Likert scale items, including MIV, MSC, and MUV, are treated as continuous variables. The rationale for adopting this assumption, along with its advantages and disadvantages, are extensively discussed in the estimation section (3.4.2. Estimation) of the method chapter.

Japan is the country with the highest mean achievement score of approximately 594, while English students had approximately 517 achievement scores, and the average for Turkish students had around 495. The CenterPoint scale provided by TIMSS is 500 as the reference point for country comparisons.

Türkiye's mathematics achievement mean is slightly below the reference point (500), while England's is slightly above it. On the other hand, Japan has an average mathematics achievement score well above the reference point and is the fourth highest scoring country in the overall ranking. It is notable that, while the mathematics scores of boys and girls are very similar in Japan and England, this difference is relatively greater in Türkiye (Boys = 490, Girls = 501).

The motivation factors were measured on a 4-point Likert scale, with a high score indicating high motivation. There is similar pattern between the average motivation scores across all three countries. As part of the MIMIC model section, I will evaluate the latent mean comparison between countries to determine if motivation values differ significantly. That said, according to descriptive statistics, Japanese students are apparently less motivated than English and Turkish students. English male students have the highest MSC average of 2.82, while Turkish female students have the highest MIV and MUV averages of 2.91 and 3.40, respectively. Students in all three countries have higher MUV values than MIV and MSC values, which means that extrinsic factors motivate more than intrinsic and self-concept factors. It should be noted that the effects of these values on mathematic achievement and educational aspiration may differ across countries, as will be analysed in the MIMIC model section.

Mathematics achievement scores for all countries tended towards a normal distribution; skewness values vary between -0.13 and 0.06, while kurtosis values are between -0.16 and 0.09. The skewness and kurtosis of motivation measures and HER meet the normality assumption (Kline, 2015). According to our analysis, the variable EDU ASP does not satisfy the assumption of normality. This could be due to the fact that this variable actually contains six categorical responses. The EDU ASP variable is, however, accepted as a continuous variable in this study. A robust maximum likelihood estimator (MLR) is used to control this non-normality issue in skewness and kurtosis (Muthén and Muthén, 1998–2017).

Table 4. 1 Descriptive statistics of motivational items and continuous variables

Variables	**MIV**	**MSC**	**MUV**	**HER**	**EDU ASP**	**MAT ACH**
Japan						
Skewness	.10	.34	-.48	-.05	-9.17	-.13
Kurtosis	-.62	-.54	-.25	.43	82.22	.09
Mean (SE)						
Male (N= 2,168)	2.55 (.03)	2.32 (.03)	2.95 (.03)	10.90 (.04)	4.19 (.04)	595.41 (3.58)
Female (N= 2,276)	2.32 (.03)	2.10 (.03)	2.84 (.03)	10.86 (.04)	4.28 (.04)	593.21 (3.45)
Türkiye						
Skewness	-.59	-.18	-1.41	-.14	-6.95	.06
Kurtosis	.98	.95	.78	1.46	46.31	-.16
Mean (SE)						
Male (N= 2039)	2.89 (.05)	2.65 (.05)	3.30 (.05)	9.38 (.10)	4.77 (.05)	490.18 (6.65)
Female (N= 2009)	2.91 (.05)	2.61 (.05)	3.40 (.05)	9.56 (.10)	5.19 (.04)	501.01 (5.47)
England						
Skewness	-.02	-.28	-1.14	-.02	-4.85	-.07
Kurtosis	-.90	-.76	1.00	-.01	21.58	.06
Mean (SE)						
Male (N= 1,492)	2.61 (.04)	2.82 (.05)	3.34 (.04)	10.63 (.08)	4.01 (.07)	518.24 (7.63)
Female (N= 1,715)	2.40 (.05)	2.58 (.05)	3.26 (.04)	10.81 (.09)	4.34 (.06)	515.47 (6.49)

Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement score; HER = home educational resources; EDU ASP = educational aspirations.

4.3.2. Reliability of TIMSS 2019 motivation scales

Cronbach's alpha (α) value is provided as a reliability indicator in the TIMSS 2019 technical report. In support of the α estimates, the composite reliability measure of McDonald's omega (ω) (Raykov, 2012), which is usually associated with CFA/SEM models, is also estimated in this study. Using ω provides reliability estimates directly associated with the estimated factor analysis. Because the composite reliability considers the factor loadings, error variances, and error covariances (if any), which are called method effects (e.g. the associated uniqueness (errors) associated with negatively worded items), seems to be more precise than the estimates provided by α (Brown, 2015; Raykov, 2012). The cut-off criteria of McDonald's omega are the same as α; 0.600–0.700 is considered acceptable in exploratory research (Hair et al., 2019).

In response to research question 1a, I calculated McDonald's omega (ω) values to examine TIMSS 2019 composite reliability scores (see Table 4.2). In addition, the alpha values in the TIMSS 2019 technical report are given in Table 4.2. In order to examine whether the negatively worded items also influence reliability, I calculate McDonald's omega (ω) separately for the situation of the error terms of negative items in correlated and uncorrelated.

The results indicate that the composite reliability assumption is provided for alpha and omega values in all cases (>0.700). Overall sample reliabilities (ω: 0.854–0.931) are of a desirable standard for the ω estimates (Hair et al., 2019). As in the previous section, negative coded items affected reliability, at the least in England and the most in Türkiye. The results showed that reliability appeared to be high without accounting for negative items' error terms correlation (ω), but low reliability emerged after the model was appropriately estimated (ω_i). There is almost no difference between

McDonald's omega (ω) and Cronbach's alpha (α) estimates for the factors without negative items error terms correlated for all three factors. Nevertheless, McDonald's omega estimates with correlated error terms (ω_i) are slightly lower than corresponded omega without correlated error terms (ω). Thus, calculating alpha and omega without error terms leads to overestimating construct reliability.

Table 4. 2 Composite reliability of TIMSS 2019 motivational constructs used in the study

Country	MIV			MSC			MUV	
	ω	**ω_i**	**α**	**ω_i**	**ω**	**α**	**ω**	**α**
Japan	0.94	0.90	0.94	0.88	0.79	0.90	0.85	0.87
Türkiye	0.92	0.84	0.92	0.86	0.73	0.89	0.85	0.88
England	0.93	0.87	0.93	0.86	0.72	0.88	0.87	0.88
Total	0.93	0.85	-	0.854	0.71	-	0.85	-

Note. Mathematics intrinsic value (MIV); mathematics self-concept (MSC); mathematics utility value (MUV); α: Cronbach alpha (derived from TIMSS 2019 technical report (Yin and Fishbein, 2019); ω : Composite reliability (CR), ω_i : Composite reliability with correlated uniqueness.

4.3.3. Factor structure of TIMSS 2019 mathematics motivation scales

This section examines the psychometric properties of TIMSS 2019 motivation scales used in this study under the subheadings of the goodness of model fit statistics and method effects, factor loading of TIMSS 2019 motivation scales, and measurement invariance.

4.3.3.1. The goodness-of-fit statistics and method effects

This study examines TIMSS motivational factors within the scope of the expectation-value theory. The justification and clarification of the theoretical approach have been detailed in the theoretical framework and literature chapters. In the TIMSS 2019 study, student motivation factors are given under three substructures: "students confident in mathematics", "students like learning mathematics", and "students value mathematics". Although TIMSS does not specify a clear theoretical basis for the development of items for these factors, self-determination theory (Deci and Ryan, 1985) and self-concept theory (Marsh and Craven, 2006) are mentioned in the assessment framework section (Hooper et al., 2017). In this section, I examine models 1, 2, 3, 4, 5, 6, and 7, which are detailed in Table 4.3. All the data were analysed as a single group in the first four models and in the others multi-group models applied. It is the purpose of this section to address the second hypothesis (H2) to determine whether the current motivational items used in the study support TIMSS's a priori factor structure. Additionally, I conducted a comparative analysis to determine whether the ESEM measurement model provides a better measurement model than the CFA model (H3). Moreover, the effect of negative items on the model fit was examined to test the fourth hypothesis (H4). The analysis starts with a base model based on the a priori factor structure of TIMSS motivation measures in the TIMSS assessment framework.

Models 1 and 3 are the basic CFA and ESEM models that reflect the a priori factor structure of motivational factors. The a priori model (Table 4.3) posits that 23 motivation items can be explained by three factors, namely MIV, MSC, and MUV. This result confirms our second hypothesis (H2). Although the model fit indices of this model do not fit the data adequately for Models 1 and 3, this result is consistent with our fourth hypothesis. According to hypothesis three (H4), negatively worded items in the questionnaire reduce model fit statistics, as negative wording is perceived as a potential source of variance unrelated to the construct being measured, especially in studies involving young individuals.

In this study, to test the method effect, a total of six negatively worded items' errors – four items on the MSC scale, and two on the MIV – are correlated with each other. In model M1 the entire data set is run as a single group, and it was observed that the model fit indexes (i.e. CFI = 0.901, TLI = 0.889, and RMSEA = 0.066) almost meet the acceptable level fit apart from TLI (i.e. CFI >0.90, TLI >0.90, and RMSEA <0.08). However, in Model M2, when the error terms of the negatively coded items freely covary with each other, the goodness of model fit has significantly increased reaching an adequate level (i.e. CFI = 0.946, TLI = 0.937, and RMSEA = 0.050).

Similar to M1 and M2, I first estimate the ESEM model (M3) without correlated uniqueness and then add correlated uniqueness in M4. Like the CFA models, the goodness-of-fit statistics are substantively increased in M4 (i.e. CFI = 0.953, TLI = 0.941, and RMSEA = 0.048) compared to M3 (i.e. CFI = 0.908, TLI = 0.890, and RMSEA = 0.065). These findings validate the hypothesis that negatively worded items should be correlated with each other to improve and reach an acceptable model fit. In other words, if the negative and positively worded items in the data set are used together, these items' error terms should correlate to reduce the method effect to achieve an acceptable model fit. Other studies in the literature also support this result (Bofah and Hannula, 2015; Chiu, 2012; Marsh et al., 2013a).

The total groups' analysis results (compare model M1 vs M3 and M2 vs M4) also show that ESEM resulted in a higher level of fit to the data than CFA (lower information criteria and RMSEA and changes in CFI/TLI ≥.010). In contrast to CFA, in the ESEM approach, cross-loading of items is allowed but targeted to be close to zero. With this aspect, the ESEM approach offers the advantages of both CFA and EFA factor analysis in a single solution. A comparison of the CFA and ESEM models is also conducted on the multigroup analyses. Model 5 and Model 7 model fit results indicate that the ESEM model (M7) is significantly better than the CFA model (M5) as the difference between CFI and TLI is greater than 0.01 between the models (for M5 CFI: 0.942, TLI: 0.932, RMSEA: 0.052; for M7 CFI: 0.952, TLI: 0.941, RMSEA: 0.049). This result is consistent with the third hypothesis (H3), which means that it is more appropriate to use the ESEM model with the current TIMSS 2019 data. Due to this, ESEM models are applied in subsequent analyses. The characteristics and factor loadings of these items in the ESEM and CFA analyses were also examined in the following section based on the metric invariance analysis of models M6 and M8. In addition, Models M7–M11 also meet the criteria for an adequate model fit. The results of CFA-based M6 and ESEM-based M8 are compared in detail in the following section in order to evaluate the differences between factor loadings in the two models. Moreover, Models M7–M10 examine the measurement invariance. Finally, a correlation analysis based on the MIMIC model is used to explore the relationship between motivational constructs, educational outcomes, and demographic indicators in model M11. The Mplus syntaxes of the 11 models developed for this chapter are presented in appendix 3.

Table 4. 3 Summary of goodness-of-fit statistics for CFA and ESEM models

Model	**Factors in the model**	**MLR $\chi 2$**	**df**	**CFI**	**TLI**	**RMSEA**	**Model Description**
Total Group Analysis							
M1	MIV + MSC + MUV	11,625.015	227	0.901	0.889	0.066	Total group CFA model without negative item correlated
M2	MIV + MSC + MUV, CU	6,455.818	217	0.946	0.937	0.050	Total group CFA model with negative item correlated
M3	MIV + MSC + MUV	10,804.654	213	0.908	0.890	0.065	Total group ESEM model without negative item correlated
M4	MIV + MSC + MUV, CU	5,597.824	203	0.953	0.941	0.048	Total group ESEM model with negative item correlated
Multigroup Analysis							
M5	MIV + MSC + MUV, CU, No inv	7,480.838	651	0.942	0.932	0.052	All parameters are freely estimated in CFA model across groups
M6	MIV + MSC + MUV, CU, Metric (FL inv)	8,363.693	691	0.933	0.928	0.053	Factor loadings are held equal (invariant) in CFA model across the groups
M7	MIV + MSC + MUV, CU, No inv	6,221.931	609	0.952	0.941	0.049	All parameters are freely estimated in ESEM model across groups
M8	MIV + MSC + MUV, CU, Metric (FL inv)	7,332.636	677	0.943	0.936	0.050	Factor loadings are held equal (invariant) in ESEM across the groups
M9	MIV + MSC + MUV, CU, Scalar (FL and item intercept inv)	13,576.309	717	0.890	0.884	0.068	Factor loadings and item intercepts are held equal in ESEM across the groups

M10	MIV + MSC + MUV, CU, Partial Scalar	8,569.051	699	0.933	0.927	0.054	Eight item intercepts are freely estimated
Extended MIMIC model							
M11	MIV+MSC+MUV+EDU ASP+MACH+HER+GENDER+ SCHOOL SES, CU	10,555.038	939	0.925	0.916	0.051	Motivation factors are related to the educational outcomes and demographic variables in ESEM model

Note. These are average results over five imputed data. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement score; HER = home educational resources; EDU ASP = educational aspirations. MLR = robust maximum likelihood estimator, χ^2 = chi-square, df = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, FL = factor loadings, CU = correlated uniqueness. All models are based on a 5-imputed data set. MIVN2, MIVN3, MIV9, MSC1, MSC6, MSCN8, MUV2, MUV8 item intercepts freely estimated for both partial CFA and ESEM models. Values below the cut-off point (i.e. CFI >0.90, TLI >0.90 and RMSEA <0.08) are highlighted in red.

4.3.3.2. Factor loadings of TIMSS 2019 motivation scales

This section examines the factor loading of motivation measures using both CFA and ESEM approaches. First of all, it should be noted that only cross-loading between MIV and MUV factors is allowed in ESEM models since task value and MSC are clearly separate theoretical concepts. In this section, factor loadings of TIMSS 2019 motivational constructs (MIV, MUV and MSC) in CFA and ESEM models are presented in Table 4.4 according to the results of model 6 and model 8. Research question 1b (Comparison of CFA and ESEM) is evaluated through the model fit value in the previous section; this section examines and compares factor loadings of latent variables in CFA and ESEM models. Therefore, this section is a continuation of research question 1b.

The analysis of parameter estimates shows that both the CFA and ESEM models are capable of providing reasonable factor loadings. With respect to ESEM, the 23 target factor loadings are within a preferred range as shown in Table 4.4. The items' factor loadings ranged from 0.448 to 0.935 for ESEM, and 0.398 to 0.904 for CFA. The non-target loadings for the ESEM solution are systematically lower (from 0.001 to 0.159) than the target loadings for the ESEM solution (non-target loadings are constrained to be zero in the CFA solution). CFA and ESEM solutions appear to follow very similar patterns when both target and non-target factor loadings are considered. On the other hand, a detailed analysis of the cross-loadings indicated that the ESEM model provides a more flexible solution psychometrically. An example of this is MIVN2 (main-loading = 0.513; cross-loading = 0.140), which shows a substantial loading on the MIV while exhibiting moderate loading on the MUV, which is not the target scale. There are also noticeable cross-loadings on MIV for MUV1 (main-loading = 0.465; cross-loading = 0.159) and MUV2 (main-loading = 0.474; cross-loading = 0.102).

These results appear to indicate that multiple meaningful cross-loadings exist and that this overlap in conceptual ideas should not be ignored, as in the CFA model. A consideration of cross-loadings is important because the omission of even a few small cross-loadings might result in biased correlations (Morin et al., 2013).

Table 4. 4 Unstandardised factor loadings (λ) for confirmatory factor analysis (CFA) and exploratory structural equation modelling (ESEM)

Item Wording	**ESEM (model 8)**			**CFA (model 6)**		
	MIV(λ)	**MUV(λ)**	**MSC(λ)**	**MIV(λ)**	**MUV(λ)**	**MSC(λ)**
I enjoy learning mathematics	**0.768**	-0.022		0.758		
I wish I did not have to study mathematics	**0.513**	0.140		0.589		
Mathematics is boring	**0.612**	0.040		0.634		
I learn many interesting things in mathematics	**0.569**	0.091		0.615		
I like mathematics	**0.911**	-0.052		0.885		
I like any schoolwork that involves numbers	**0.676**	0.017		0.686		
I like to solve mathematics problems	**0.823**	-0.019		0.816		
I look forward to mathematics class	**0.738**	0.001		0.739		
Mathematics is one of my favourite subjects	**0.935**	-0.062		0.904		
I think learning mathematics will help me in my daily life	0.159	**0.465**			0.557	
I need mathematics to learn other school subjects	0.102	**0.474**			0.534	
I need to do well in mathematics to get the job I want	-0.007	**0.601**			0.593	

It is important to learn about mathematics to get ahead in the world	-0.009	**0.705**		0.698	
Learning mathematics will give me more job opportunities when I am an adult	-0.074	**0.675**		0.621	
My parents think that it is important that I do well in mathematics	-0.073	**0.448**		0.398	
It is important to do well in mathematics	-0.023	**0.499**		0.480	
I usually do well in mathematics			**0.664**		0.665
Mathematics is more difficult for me than for many of my classmates			**0.462**		0.461
Mathematics is not one of my strengths			**0.678**		0.676
I learn things quickly in mathematics			**0.641**		0.640
I am good at working out difficult mathematics problems			**0.692**		0.692
My teacher tells me I am good at mathematics			**0.581**		0.581
Mathematics is harder for me than any other subject			**0.578**		0.575

Note. The results are based on model M6 and M8 (metric invariance models) and are averaged over five imputed data sets. Factor loadings are unstandardised estimates. Factor loadings were constrained to be equal across three countries. For model identification, factor variances are fixed to one in CFA model. All estimates are statistically significant at (p <.05). Items identified with a star (*) are all negatively worded items. All items were reverse coded, so that higher values correspond to higher responses. Boldface indicates target ESEM factor loadings.

4.3.3.3. Measurement invariance

This section analyses research question 1e and hypothesis 5. This is done using model 5 through model 10 in table 4.3.

Multigroup CFA is used to test to what extent the factor structures of MIV, MSC, and MUV can be generalised to each group, and it is not affected across the countries. As mentioned in the data and methods chapter, measurement invariance analysis usually begins with a configural invariance. Configural invariance refers to the number of factors and the loading patterns that are the same across the groups. In other words, the specific items loaded on each of the relevant factors are the same for each group (Rhudy et al., 2020). As seen in Table 4.3, the configural invariance model for both the CFA model (M5) and the ESEM model (M7) fitted the data well (i.e. for M5 CFI = 0.942, TLI = 0.932, and RMSEA = 0.050; for M7 CFI = 0.952, TLI = 0.941, and RMSEA = 0.049) and showed support for configural validity across nations, and thus allowed the passing to test metric invariance.

Metric invariance (also called weak invariance) means that the same items are loaded on the same factors for each group and that the actual magnitude of the loads between groups for each relevant item is the same. To achieve metric invariance, the decrease in CFI and TLI between the configural and the metric model should be less than 0.01 for CFI and TLI, while the increase in RMSEA should not be more than 0.015 (Kline, 2005). The fit statistics of the metric invariance model CFA model (M6) is CFI = 0.933, TLI = 0.928, and RMSEA = 0.058, while CFI = 0.943, TLI = 0.936, and RMSEA = 0.050 for ESEM (M8). Therefore, it can be concluded that metric invariance is achieved, and the data fit well for both CFA and ESEM. The MIV, MSC, and MUV variables are invariant in these three countries. In other words, these motivational structures, which have the same dimensional measurement characteristics, can be generalised for Japan, Türkiye, and England.

After that, I test the scalar invariance. Also referred to as strong invariance, this imposes the same constraints as configural and metric invariance (equal factor loadings) but keeps the item intercepts equal across groups. For the scalar invariance model test, I look at the change in the model fit indexes similar procedure to the metric invariance test. However, the scalar invariance model M9 fit is not as good as in the metric model. Since the model failed to accomplish cut-off criteria (less than 0.90 for CFI and TLI) and the drop in CFI and TLI is greater than 0.01, scalar invariance is not achieved. Chen (2008) has explained this situation as follows: due to possible cultural differences and social desirability, measurement invariance, especially full scalar invariance, is rarely achieved in cross-cultural studies. As scalar invariance is essential to make valid interpretations and comparison of the latent means, the alternatively partial scalar model should be tested to compare latent means across the groups (Steenkamp and Baumgartner, 1998). Following the suggestion of Byrne et al. (1989), I presented the partial invariance model by freely estimating the intercepts of items with the most variants in modification indices. Accordingly, model 10 (M10) freely estimates four of seven “mathematics self-concept items”, three of nine “mathematical value items”, and one of seven “mathematical utility value items” across countries. The result of model M10 supports partial invariance for item intercepts, allowing comparison of latent mean differences between countries (CFI = 0.933, TLI = 0.927, and RMSEA = 0.054). As full metric invariance and partial scalar invariance is achieved across the countries, any estimated latent mean differences between countries can be reliably predicted (Chen, 2008; Steenkamp and Baumgartner, 1998).

4.3.4. Construct Validity of Mathematics Scales: Relations to Correlates via MIMIC model

This section consists of two sub-sections: "latent mean differences" and "MIMIC model: the relationship among motivational constructs, background variables, and educational outcomes". This section studies research questions 2a and 2b and related hypotheses 6 and 7. For this purpose, model 11 is developed and the necessary information is presented in table 4.5.

4.3.4.1. Latent mean differences

After applying scalar invariance in which the intercepts of measurements are kept constant between groups, the overall mean structure of the factor model between groups is fixed to 0 in one group and can be defined by freely estimating the factor mean in all other groups. A positive value indicates that the compared groups had higher latent mean values than the reference groups; a negative value indicates the opposite. This analysis considered the Japanese sample as the reference group and therefore fixed the latent means of the three constructs to 0 in the Japanese sample in order to assess the size and direction of differences between the remaining four countries.

As an illustration of the rationale for these comparisons, the first discussion focuses on the latent mean of the MIV in the Japanese sample which, although fixed to 0, has a standardised latent mean value of 0.764 in the Turkish sample and 0.257 in the English sample (Table 4.5). Therefore, the mean of MIV in Japan is significantly lower than in Türkiye and England, and since these are standardised mean differences (SDs of latent variables are 1.0 in all countries), the difference between countries is in units of standard deviation. A similar pattern is seen for MSC and MUV latent constructs. It is possible to say that Turkish students have a higher motivation value than English and Japanese students and that the latter are the least motivated across the countries. This result supports our hypothesis that Japanese students have lower motivation than Turkish and English students (H6). In the following section, I examine the correlations between motivational constructs (MSC, MIV and MUV) and background factors (gender and HER) and educational outcomes (mathematics achievement and educational aspirations).

Table 4. 5 Estimated latent means and correlations (Model 11)

Country/ Variable	MIV	MSC	MUV	HER	EDU ASP	MAT ACH	Gender
Japan							
Latent Means (SE)	0.000	0.000	0.000	8.195 (.11)	3.485 (.07)	7.063 (.15)	-
MIV	1.000						
MSC	.812	1.000					
MUV	.460	.350	1.000				
HER	.141	.205	.128	1.000			
EDU ASP	.216	.274	.213	.306	1.000		
MAT ACH	.433	.583	.230	.352	.447	1.000	
Gender	.161	.203	.077	.013[a]	-.036[a]	.013[a]	1.000
Türkiye							
Latent Means (SE)	.764 (.05)	.801 (.04)	.714 (.05)	5.175 (.14)	4.159 (.14)	4.527 (.13)	-
MIV	1.000						
MSC	.804	1.000					
MUV	.624	.539	1.000				
HER	-.006[a]	.134	.065	1.000			

EDU ASP	.206	.266	.274	.335	1.000		
MAT ACH	.251	.460	.244	.480	.469	1.000	
Gender	-.007[a]	.041	-.074	-.048	-.173	-.050	1.000
England							
Latent Means (SE)	.257 (.04)	.681 (.05)	.626 (.04)	7.180 (.15)	2.964 (.07)	5.766 (.24)	
MIV	1.000						
MSC	.749	1.000					
MUV	.487	.391	1.000				
HER	.041[a]	.097	0.040[a]	1.000			
EDU ASP	.189	.206	.226	.310	1.000		
MAT ACH	.242	.428	.099	.410	.371	1.000	
Gender	.127	.198	.078	-.060[a]	-.121	-.015[a]	1.000

Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement score; HER = home educational resources; EDU ASP = educational aspirations. Non-significant values are marked with "a". 1= Female, 2 = Male. SE = Standard Errors

4.3.4.2. MIMIC model the relationship among motivational constructs, background variables, and educational outcomes

The first step is to evaluate the correlations between the three multi-item motivation factors (see Table 4.5). Although correlation patterns are similar in each country, the correlations between MIV and MSC are always highly correlated (0.749–0.812), which potentially undermines their discriminant validity.

In all three countries, a positive correlation between motivational constructs and achievement was found, but the effect size for MUV and achievement in England is small. As shown in Table 4.5, the achievement is strongly correlated with MSC, while correlations with MUV (0.230 for Japan, 0.244 for Türkiye, and 0.099 for England) and MIV (0.433 for Japan, 0.251 for Türkiye, and 0.242 for England) are smaller.

Students’ long-term educational aspirations (EDU ASP) are positively related to the motivational constructs in Japan, Türkiye, and England. The interesting point here is that while MSC is more associated with mathematics achievement in all three countries, MUV has a stronger relationship with EDU ASP than other motivational structures. In Türkiye and England, there is a statistically no relationship between HER and the MIV, whereas in Japan, a positive correlation exists. In addition, there is a significant positive correlation between HER and the student’s self-concept (MSC) for all three countries.

It is particularly concerning that the correlation between the MIV and the MSC is very high in the TIMSS data. There is a robust correlation between MSC and MIV for all countries, which could pose a problem for discrimination validity. However, the relationship between these and external variables reveals distinct structures and ensures discriminant validity. In all three countries, the MSC variable appears to be more strongly correlated with mathematics achievement than MIV. Moreover, the MIV variable is not significantly associated with HER in Türkiye and England and has a stronger correlation in Japan.

In England and Japan, the relationship between motivation factors and gender is higher in favour of male students. In particular, the MSC correlation of male students is higher than female students (Japan = 0.203; England = 0.198). However, no relationship is found between mathematic achievement and gender in both countries. On the other hand, among Turkish students, there is a significant but small relationship between gender and motivation structures in favour of male students only with MSC, while there is no relationship with MIV and MUV. There was also a weak correlation between achievement and gender in favour of Turkish female students.

4.4. Summary

A psychometric evaluation of the motivation measurements in TIMSS 2019 is presented in this chapter. Although some of the findings of our study are consistent with those of the literature, there are other findings which are not. The results are organised into three subheadings: The subsections begin with a reliability analysis and proceed to examine the “Factor Structure of TIMSS 2019 Mathematics Motivation Scales” finally, latent mean and correlation analyses are conducted to evaluate construct validity.

In the reliability analysis, MC Donald’s omega was calculated in addition to the alpha value provided in the TIMSS 2019 technical report (Yin and Fishbein, 2019). In short, some studies in the literature have found that TIMSS motivation measures are unreliable in non-Western countries (Bofah and Hannula, 2015; Marsh et al., 2013a). However, the results of this study, contrary to the literature, provided strong evidence that TIMSS motivation measures are reliable measures based on alpha and omega estimation in Türkiye, England, and Japan. Additionally, the omega value was estimated in this study by considering the effect of negatively worded items. In comparison to alpha and omega that ignore negative items’ effects, the estimation is lower but acceptable. Therefore, the omega reliability test may be preferred to avoid overestimation, which is an issue for alpha in a structure with latent variables used and negative items present in this study.

In the second part of the results, the factor structures of TIMSS 2019 motivation items, including model fit, negative items’ effect, factor loadings, and measurement invariance are examined. As a first step, the CFA and ESEM measurement models were compared, and the ESEM measurement model provided a better model fit consistent with the literature and our expectations. In a second analysis, we examined the effect of negative items on the model fit, and we found that the correlation of the error terms between negative items significantly increased the fit, as expected. In addition, factor loadings were examined in the CFA and the ESEM models and, as a result, it was demonstrated that the ESEM model has an advantage over the CFA model when it comes to the possibility of cross-loading in structures that are theoretically highly correlated. The final analysis investigated whether the motivational structures of TIMSS 2019 provide measurement invariance in Japan, Türkiye, and England. According to the results of this analysis, configural and metric measurement invariance were provided, but scalar invariance could not be achieved. Based on the literature, the intercepts of some items were released, and the measurement invariance was obtained by applying the partial scalar invariance test.

In the final results of this section, in addition to the motivation factors, variables such as gender, HER, mathematics achievement, and educational aspirations were added to the model. First, the latent means of motivation values of Japanese, Turkish, and English students were compared. In line with our expectations, Japanese students have the lowest motivation among the three. The second part of this section examined the relationship between motivational factors and other variables. According to the results of this model, a high correlation was found between MIV and MSC in all three countries, which poses a problem in terms of discriminant validity. However, when the relationship of MSC and MIV with achievement and other variables was examined, evidence was obtained that these structures support discriminant validity. For example, MSC and achievement have a stronger relationship than MIV, and MSC has a significant relationship with HER, while MIV does not.

Overall, this chapter focuses on TIMSS 2019 motivational constructs from a psychometric perspective, while the next chapter examines motivational constructs in relation to mathematics and educational aspirations.

Chapter 5: The Relationship Between Expectancy-values and Their Interactions with Educational Outcomes

5.1. Introduction

This chapter studies the effects of motivation on educational outcomes within the framework of expectancy-value theory (EVT) and using multi-group structural equation models (SEM). This section also examines the latent interactions between expectancy and value in predicting educational outcomes with the unconstrained product indicator approach under the SEM (see chapter 3). The aim of the chapter is to investigate its power to predict the educational outcomes of motivation factors and the interaction effect of expectancy (self-concept) and value beliefs (intrinsic and utility value) that have not been used recently in the application of EVT (Nagengast et al., 2011) within its scope in Japan, Türkiye, and England.

This chapter develops as follows: First, the relevant research questions and hypotheses are introduced. Next, the analysis of SEM is explained. The results of the analysis are then presented in the following section. This section provides the results of SEM models for each outcome variable (mathematics achievement and educational aspirations) separately. Finally, a summary of the chapter is presented.

5.2. Research Questions and Hypotheses

In this chapter, I focus on research questions 3a, 3b, and 3c.

(3) What is the relationship between student motivation and educational outcomes (mathematics achievement and educational aspirations)?

  a) How well do motivational factors predict educational outcomes?

H8: Value beliefs will have a positive effect on educational outcomes. However, if expectancy beliefs are incorporated into the model, value beliefs will weaken, and expectancy beliefs (MSC) become the strongest predictors for mathematics achievement. However, educational aspirations will remain more strongly influenced by value beliefs than expectancy beliefs.

b) Is there an interaction effect between expectancy beliefs and value beliefs on educational outcomes?

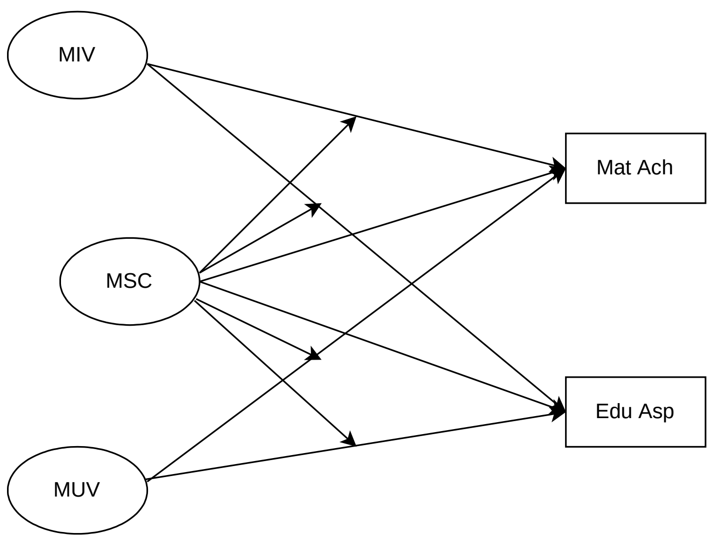
H9: Self-concept and value will have a significant interaction effect on mathematics achievement and educational aspirations.

5.3. Analysis of Structural Equation Models

In order to answer research questions 3a and 3b, I use SEM to predict mathematics achievement and educational aspirations and test the latent interaction effect of expectancy and value beliefs. The hypothesised model is shown in Figure 5.1, which shows the arrows from MIV, MSC, and MUV to outcome variables (MAT ACH, EDU ASP) representing the direct effect, while the arrows following the path from MSC to MIV and MUV represent the interaction between MSC and MIV/MUV.

First, I specify a set of models with an increasing number of predictors in the SEM. In the first model (M12), I use value variables MIV and MUV as predictors. In the next step, the expectancy variable (MSC) is added to the model (M13) along with task value variables to predict the educational outcomes (mathematics achievement and educational aspirations). In the next step, I use the unconstrained approach to test the latent interaction between expectancy and value beliefs to predict mathematics achievement and educational aspirations in Models 14 and 15. Although the latent variable model with interactions became available in theory in the 1980s (Kenny and Judd, 1984), SEMs with latent interactions have relatively recently become more accessible to applied researchers (Klein and Moosbrugger, 2000; Marsh et al., 2004). Latent interaction terms (product terms) are created by multiplying mean-centred indicators of first-order effects (i.e. MSC x MIV/MUV). As suggested by Marsh et al. (2004), interaction terms are formed using a matched-pair strategy. Based on this strategy (Marsh et al., 2004), the indicator should be used once to create a product term, and indicators are matched based on their factor loadings (i.e. the best item in MSC with the best item in MIV). Due to the different number of indicators for the latent variables (MSC = 7; MIV = 9), two MIV indicators are excluded from the product terms of MSCxMIV but kept in MIV. The results of the SEM analyses for mathematics achievement and educational aspirations are presented in Tables 5.1 and 5.2, respectively. The significant interaction effect is also demonstrated with graphics.

Figure 5. 1 The hypothesised conceptual framework of EVT factors and their interaction in relation to mathematics achievement and educational aspirations



Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement; EDU ASP = educational aspirations. The arrow from MSC to the path of MIV and MUV to the outcome variables represents the interaction.

5.4. Results

A total of four models were developed to answer research Question 3 (M12 to M15). These models use mathematics achievement and educational aspiration variables as dependent variables. However, the results are reported in two separate tables for each dependent variable for clarity. The section, therefore, consist of two main subsections: the first section presents the results related to mathematics achievement, and the second presents the results related to educational aspiration.

5.4.1 Predicting mathematics achievement

This section discusses regression models (Models 12, 13, 14, and 15) developed using the SEM approach. Models 12 and 13 examine the additive effects of expectancy-value variables, while Models 14 and 15 emphasise their multiplicative (interaction) effects on the dependent variables. A summary of the statistical parameters for each of these models can be found in Table 5.1.

I first specify Model 12, which includes only value beliefs variables (MIV and MUV). The SEM provided a good fit for the data: RMSEA: 0.57, CFI: 0.941, and TLI: 0.934. In this model, the explained variance for mathematics achievement is 0.188, 0.077, and 0.060 for Japan, Türkiye, and England, respectively. It was found that the path from MIV to mathematics achievement is stronger than the corresponding path from MUV. Specifically, the predictive β power of the path from MIV to mathematics achievement is 0.414, 0.165, and 0.256 for Japan, Türkiye, and England, respectively. In addition, MUV has a small effect size for Japan ($\beta = 0.040$) and is not statistically significant for England ($\beta = -0.025$). Nevertheless, the predictive power of the MUV ($\beta = 0.142$) for Türkiye is almost comparable to the MIV ($\beta = 0.165$). In the next step, I regress mathematics achievement on MSC along with MIV and MUV (see Model

M13 in Table 5.1). Statistical analysis of the data indicates that the model fits the data reasonably well: RMSEA = 0.51, CFI = 0.935, and TLI = 0.927. The results show that MSC is statistically significant for the prediction of mathematics achievement across all three countries (β = 0.702 (Japan), 0.700 (Türkiye), 0.587 (England)) with an explained variance increase to 0.367, 0.244, and 0.214 for Japan, Türkiye, and England, respectively. The important point in this model is that the MIV variable loses its predictive power when the MSC variable is added to the model. Possibly, this is caused by the high correlation between the two variables, as revealed by the correlation analysis conducted in the previous chapter (see Chapter 4, Section 4.3.4.2). In other words, this high correlation between MSC and MIV likely causes a suppression effect (Meyer et al., 2019). The result of this model is consistent with the literature-based hypothesis (H8) we have developed (Guo et al., 2015; Lauermann et al., 2017; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012) in that a high correlation between intrinsic value and self-concept makes intrinsic value less predictive when self-concept is included in the analysis and it becomes the strongest predictor.

Table 5. 1 Structural equation model results for predicting mathematics achievement with expectancy and value beliefs and their interaction

	M12			M13			M14			M15		
	JPN	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**
Value Beliefs												
Mathematics Intrinsic Value (MIV)	.414 (.019)	.165 (.031)	.256 (.033)	-.152 (.041)	-.396 (.041)	-.163 (.045)	-.140 (.039)	-.310 (.051)	-.185 (.048)	-.152 (.040)	-.414 (.042)	-.156 (.045)
Mathematics Utility Value (MUV)	.040 (.020)	.142 (.026)	-.025[a] (.032)	.061 (.020)	.097 (.024)	-.044 (.027)	.060 (.019)	.090 (.025)	-.042[a] (.027)	.043 (.020)	.207 (.030)	-.059 (.030)
Expectancy Beliefs												
Mathematics Self-Concept (MSC)				.702 (.042)	.700 (.039)	.587 (.041)	.711 (.042)	.696 (.042)	.614 (.045)	.721 (.042)	.660 (.043)	.586 (.040)
Expectancy-Value Interaction												
MSCxMIV							-.075 (.018)	.190 (.025)	.080 (.027)			
MSCxMUV										-.068 (.019)	.143 (.027)	-.029[a] (.022)

R^2	.188 (.015)	.077 (.011)	.060 (.013)	.367 (.023)	.244 (.021)	.214 (.024)	.368 (.023)	.275 (.020)	.218 (.024)	.373 (.022)	.254 (.021)	.214 (.023)
Model Fit												
X2	5,454.733			8,755.942			11,755.647			11,254.87		
df	404			799			1,325			1,325		
RMSEA	.057			.051			.046			.047		
CFI	.941			.935			.913			.900		
TLI	.934			.927			.905			.891		

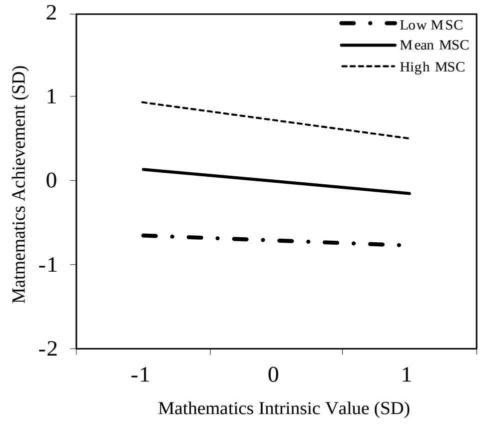
Note. Motivation factors are modelled as latent variables. Standardised regression coefficients are given, with standard errors in parentheses. X2 = chi-square; df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker-Lewis index. a represents non-significant associations.

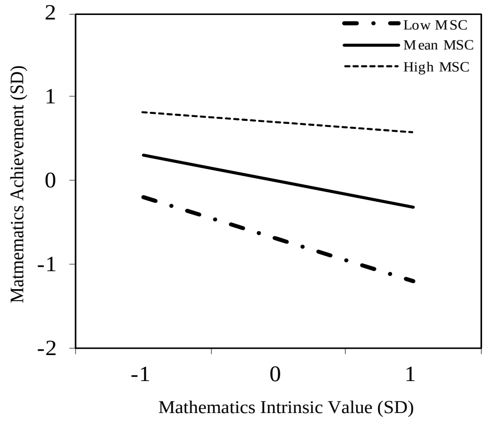
In addition, one of the major contributions of this section is the testing of expectancy and value interaction. To do this, Models 14 and 15 were developed. In Model 14, the interaction term named MSCxMIV of expectancy (MSC) and intrinsic value (MIV) value has been added to the model. Although there is a clear decrease compared to the previous model (M13) the goodness-of-fit statistics indicate that the model fits the data well: RMSEA = 0.46, CFI = 0.913, and TLI = 0.905. This model indicates that the MSCxMIV variable has a statistically significant effect across all three countries (β = -0.075 for Japan; β = 0.190 for Türkiye; β = 0.080 for England). There is also an increase in the explained variance rates. As compared to Model 13, the variance explained increased from 0.367 to 0.368 for Japan, from 0.244 to 0.275 for Türkiye, and from 0.214 to 0.218 for England in Model 14. Thus, it can be seen in this model that the explained variance values for Japan and England have increased at a small rate, while an increase of more significance is observed in Türkiye.

The significant interaction effect is presented with a simple slope graph in Figure 5.2. Simple slopes are depicted at ±1 standard deviation from the mean of MSC and MIV. In spite of the different effect sizes of interaction terms across the three countries, the graph shows a similar trend for Türkiye and England while a slightly different pattern exists for Japan. There is a downward trend in mathematics achievement among students with low MSCs and high MIVs in England and Türkiye. In addition, for students with high MSC, high MIV does not have much effect on achievement in England and Türkiye. So, in the case of a low MSC, a high MIV does not compensate but rather has a detrimental effect on mathematics achievement for the student. On the other hand, when MIV is low, MSC contributes more to maths achievement in Japan for students with a high level of MSC. In other words, the increase in the intrinsic motivation of students with high MSC negatively affects mathematics achievement in

Japan. The possible reasons and implications for these detrimental and complex associations will be examined in the discussion section.

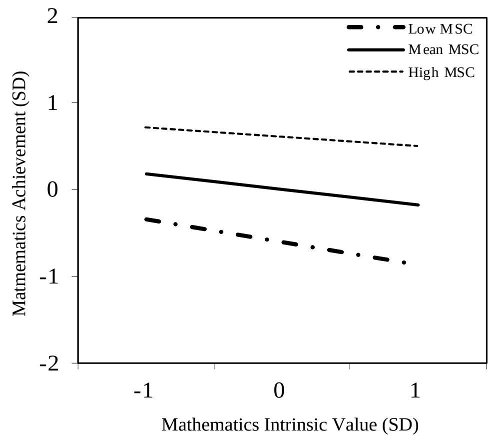
Figure 5. 2 Plots of the significant effect effects of MSC and MIV interaction on mathematics achievement.





a) Model M14 for Japan

b) Model M14 for Türkiye



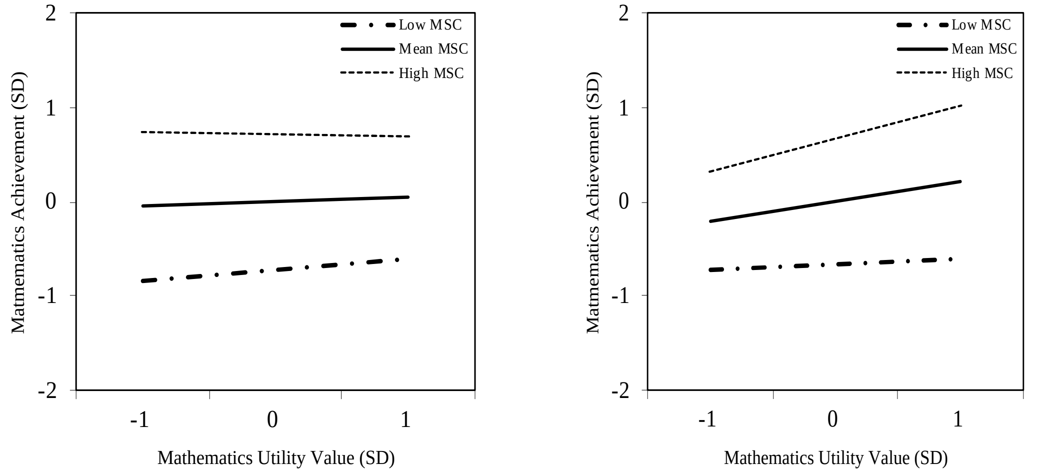
c) Model M14 for England

Note. MSC: Mathematics self-concept; MIV: Mathematics intrinsic value. Simple slopes are depicted at ±1 standard deviation from the mean of MSC and MIV

In the next model, Model 15, I tested the interaction effect of the MSC and MUV variables. The MSCxMUV variable was added to the model. The model fit values are as follows: RMSEA = 0.47, CFI = 0.900, and TLI = 0.891. It must be reminded that although there is no consensus, it is recommended that the RMSEA for model fit indicators should not be higher than 0.10, preferably less than 0.08; and CFI and TLI values should be higher than 0.90 (Hu and Bentler, 1999; Kline, 2015; Marsh et al., 2004). However, the TLI for Model 15 is 0.89. In line with Trautwein et al. (2012), I consider a model fit acceptable when at least two fit indices meet the typical range. Moreover, I also note that unique characteristics of data structures influence fit indices (Heene et al., 2011) which invalidates sweeping generalisations regarding cut-off values (Trautwein et al., 2012).

The interaction effect of MSC and MUV is statistically significant in Japan (β = -0.068) and Türkiye (β = 0.140), but insignificant in England (β = -0.029) sample (see Table 5.1). However, there is an increase in the explained variance rates for all countries when compared to the variance values explained in Model 13. The variance explained for Japan increased to 0.373, 0.254 for Türkiye, and 0.214 for England. Although the explained variance rates and effect size are seen as small, the point to be considered here is the trend that this interaction effect shows us. In Figure 5.3, I obtain a similar graph to Figure 5.2. The graph reveals that students with high mathematics utility values have higher mathematics achievement if they have high MSC beliefs. In other words, an increase only in the student's mathematical utility value does not make a positive contribution to success. Furthermore, it is evident from the graph that if students have a low MSC, the increase in MUV is of almost no significance.

Figure 5. 3 Plots of the significant effect effects of MSC and MUV interaction on mathematics achievement.



a) Model M15 for Japan

b) Model M15 for Türkiye

Note. MSC: Mathematics self-concept; MUV: Mathematics utility value Simple slopes are depicted at ±1 standard deviation from the mean of MSC and MUV

5.4.2 Predicting educational aspiration

In the same set of SEM analyses, I predict educational aspiration as an educational outcome instead of mathematics achievement. As in the previous section, I mainly aim to reveal the predictive power of expectancy-value components and their interaction terms over educational aspirations. Since the same models are used as in the previous section (predicting mathematics achievement), the values of the goodness-of-fit statistics are not discussed in this section. However, it should be borne in mind that all motivation variables are mathematics specific, while educational aspiration is a single indicator representing students' general view for further education. Given that MSC and value beliefs (MIV and MUV) are expected to measure mathematics-specific motivation, a student with a high motivation towards mathematics can contribute to educational aspirations. The summary of the statistical analysis is presented in table 5.2.

First, I regress educational aspirations on MIV and MUV in Model 12. In line with my expectation, value beliefs (MIV and MUV) are positively associated with educational aspiration when entering the model without expectancy beliefs (MSC). However, as with mathematics achievement, when MSC is added to the regression equation, the value beliefs lose their predictive powers on educational aspirations (M13). According to Model 12, the MUV is the strongest predictor of educational aspiration in Türkiye ($\beta = 0.237$) and England ($\beta = 0.176$), whereas, in Japan, MIV and MUV ($\beta = 0.145$) have almost similar effects on the educational aspiration of students. The variances explained by value beliefs are 0.063, 0.077, and 0.059 for Japan, Türkiye, and England, respectively, as shown in Model 12. It is evident from Model 13 that the effect of MIV on the dependent variable turns out to be statistically insignificant or negative, possibly due to the high correlation between MIV and mathematics achievement. As a result of the high correlation between MIV and MSC, MIV may have no unique effect on educational aspirations when expectancy and values are taken into consideration together as in model 13. Compared to model 12, the explained variance for Japan, Türkiye, and England increased from 0.063, 0.077, and 0.059 to 0.100, 0.093, and 0.071, respectively, when the MSC variable was added to model 13. This means there is a 37%, 16% and 12% increase for Japan, Türkiye, and England, respectively, in explained variance.

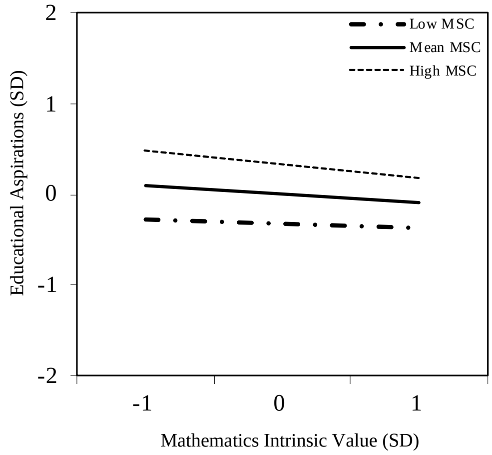
Table 5. 2 Structural equation model results for predicting educational aspiration with expectancy and value beliefs and their interaction

	M12			M13			M14			M15		
	JPN	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**	**JPN**	**TUR**	**ENG**
Value Beliefs												
Mathematics Intrinsic Value (MIV)	.149 (.017)	.058 (.027)	.103 (.024)	-.102 (.034)	-.146 (.043)	.001[a] (.035)	-.098 (.032)	-.082 (.049)	-0.016[a] (.037)	-.101 (.033)	-.154 (.043)	-.004[a] (.035)
Mathematics Utility Value (MUV)	.145 (.019)	.237 (.030)	.176 (.025)	.154 (.019)	.217 (.029)	.171 (.024)	.153 (.019)	.213 (.030)	.174 (.025)	.151 (.020)	.287 (.032)	.187 (.025)
Expectancy Beliefs												
Mathematics Self-Concept (MSC)				.314 (.032)	.254 (.037)	.145 (.036)	.325 (.032)	.247 (.040)	.168 (.037)	.317 (.033)	.225 (.042)	.146 (.036)
Expectancy-Value Interaction												
MSCxMIV							-.057 (.019)	.131 (.023)	.080 (.021)			
MSCxMUV										-.025[a] (.019)	.108 (.031)	.034[a] (.025)
R^2	.063 (.009)	.077 (.012)	.059 (.011)	.100 (011)	.093 (.012)	.071 (.012)	.101 (.011)	.109 (.012)	.078 (.013)	.106 (.011)	.095 (.011)	.074 (.012)

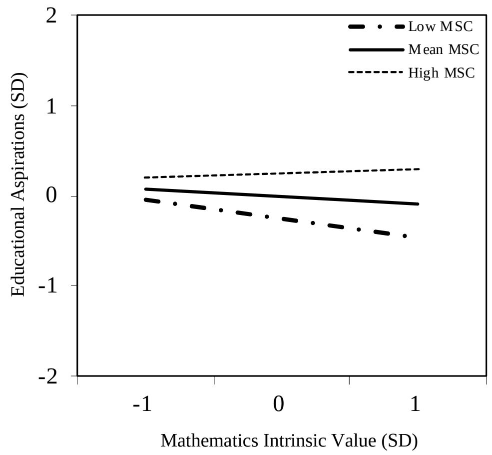
Note. Motivation factors are modelled as latent variables. Standardised regression coefficients are given, with standard errors in parentheses. X2 = chi-square; df= degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker-Lewis index. a represents non-significant associations.

The intrinsic value and self-concept interaction significantly predict educational aspirations for all countries (Figure 5.4). The graph shows a similar trend for Türkiye and England, but the direction for Japan differs. The interaction effect is evident that the effect of MSC on educational aspiration becomes more favourable when MIV is high in Türkiye and England. In other words, it has been shown that students' educational aspirations increase when their self-concepts of mathematics and intrinsic values increase simultaneously. In contrast, students with a low MSC have lower educational aspirations despite a high MIV value. The lack of MSC negatively affects education aspirations even if a high MIV is present for Türkiye and England. There is an inverse relationship between the effect of MSCxMIV interaction and educational aspirations for Japan. There is a decrease in educational aspirations among students with high MSC and MIV.

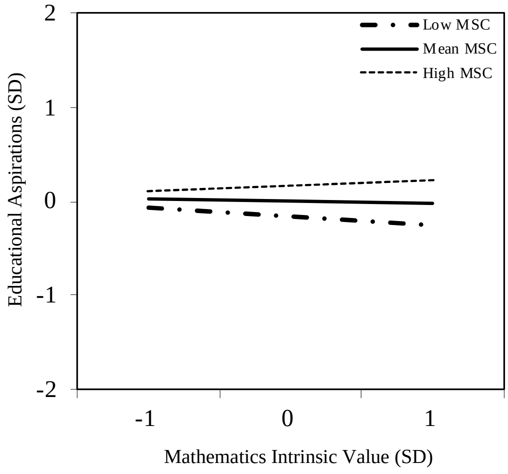
Figure 5. 4 Plots of the significant effect effects of MSC and MIV interaction on educational aspiration



a) Model M14 for Japan



b) Model M14 for Türkiye

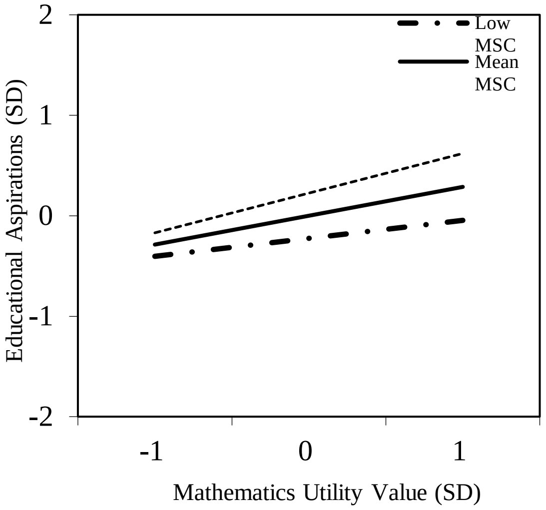


c) Model M14 for England

Note. MSC: Mathematics self-concept; MIV: Mathematics intrinsic value. Simple slopes are depicted at ±1 standard deviation from the mean of MSC and MIV

There is also a significant interaction effect between the MUV and MSC on educational aspiration only in Türkiye. Therefore, only the graph of the interaction effect of Türkiye is presented in Figure 5.5. This result suggests that MUV increases the impact of MSC on educational aspiration for students with both low and high MSC. In other words, the graph shows that an increase in MUV positively affects educational aspirations, regardless of MSC level.

Figure 5. 5 Plots of the significant effect effects of MSC and MUV interaction on educational aspiration



a) Model M15 for Türkiye

Note. MSC: Mathematics self-concept; MUV: Mathematics utility value Simple slopes are depicted at ±1 standard deviation from the mean of MSC and MUV

5.5. Summary

This chapter examines the effect of expectancy-value components on educational outcomes and their relations with each other under research Questions 3a and 3b.

The results are consistent with my expectations that show expectancy-value beliefs are significant predictors of educational outcomes for each country when they entered the model separately. Mathematics self-concept is a stronger predictor of mathematics achievement and educational aspiration than value beliefs. However, due to the strong association among expectancy-value beliefs, value beliefs lose their power when used together as predictors in the model. Therefore, it is important to bear in mind that the interaction effect must be interpreted with caution because of the high correlation between MSC and MIV. Theoretical reasons and statistical limitations will be considered further in the discussion chapter. Although there are some minor differences, the results for educational aspiration are similar to mathematics achievement. The MSC is still the strongest predictor of educational aspiration in Japan and Türkiye but not as strong as mathematics achievement. The effect size of value beliefs, especially MUV, significantly increases to predict educational aspiration compared to mathematics achievement. The MUV is the strongest predictor of educational aspiration in England. Even though MSC is a stronger predictor for Japan and Türkiye, the power of value beliefs is much closer to MSC in educational aspiration. It is consistent with my prior hypothesis that value beliefs are a better predictor of an educational aspiration than mathematics achievement.

I found mixed results regarding an interaction effect. Specifically, expectancy by intrinsic value interaction is significantly associated with mathematics achievement in all countries with similar directions but different effect sizes. Expectancy and intrinsic value interaction are also significantly associated with educational aspirations for all countries. The interaction of expectancy and utility value does not have a statistically significant effect on either mathematics achievement or educational aspirations for England. In addition, Türkiye is the only country in which expectancy and utility value interaction have statistically significant relations with educational aspirations. The results of this study reveal that given the positive effects of the interaction between expectancy and value beliefs on educational outcomes, it is essential for teachers to simultaneously emphasise increasing students' expectancy and value beliefs, with particular attention paid to strengthening the self-concept of students with low value beliefs. Although the interaction variables in this study have a small effect size, it is not that small in comparison with the variance explained by expectancy-value components. The interaction term explains approximately 10% of the variance explained by these constructs. The discussion section will explain in depth what these effects entail and their practical implications.

The next chapter will investigate the relationship between expectancy-value beliefs and demographic indicators (HER and gender) through the mediation role of expectancy and value beliefs.

Chapter 6: The Mediation Effect of Expectancy-Value Constructs on Educational Outcomes

6.1. Introduction

In Chapter 5, SEM models were used to study the additive and multiplicative effects of EVT components on educational outcomes in Japan, Türkiye, and England. This chapter seeks to explore the relationship between HER, gender and mathematics achievement, and educational aspirations with the mediating role of expectancy-value components, namely MSC, MIV, and MUV. I employed the SEM with mediation paths to estimate the structural relationship between HER, gender, MSC, MIV, MUV, and mathematics achievement and educational aspirations.

This chapter is structured in the same way as Chapter 5. First, the research questions and hypotheses provided in this chapter are presented. Then the SEM analysis is explained. Next, the results of the analysis are presented and finally, in the last section, a summary of the chapter.

6.2. Research Questions and Hypothesis

In this chapter, I focus on research Questions 4 and 5 are detailed below alongside associated hypotheses based on the reviewed literature.

(4) What is the relationship between gender mathematics achievement and educational aspirations? Do EVT factors mediate this relationship?

H10: There will be some inconsistent findings regarding gender differences in mathematics achievement and attitudes towards mathematics.

For example, several studies have demonstrated that boys are more likely to have higher values in mathematics (Marsh et al., 2005; Steinmayr and Spinath, 2008; Watt, 2004), although some have indicated no gender differences between boys and girls (Jacobs et al., 2002; Wigfield et al., 1997).

In a review of EVT-based research, Wigfield et al., (2009) provide a summary demonstrating how EVT can explain gender inequalities and achievement results in general. It has been found in multiple studies that males usually exhibit higher levels of maths self-concepts, attitudes, and effects than their female counterparts (Eccles and Wigfield, 2002; Guo et al., 2015; Marsh et al., 2013). Recently, however, cross-national meta-analyses have shown that there exist gender similarities in maths achievement at a variety of levels (Else-Quest et al., 2010). Therefore, no hypothesis was determined for this research question, and it is considered as an open question.

(5) What is the relationship between HER mathematics achievement and educational aspirations? Do EVT factors mediate this relationship?

H11: As a positive relationship between socioeconomic indicators and academic achievement is well known (Hattie, 2009; Sirin, 2005), it is expected that HER will affect both achievement and educational aspiration directly and indirectly through EVT constructs.

6.3. Analysis of Structural Equation Model

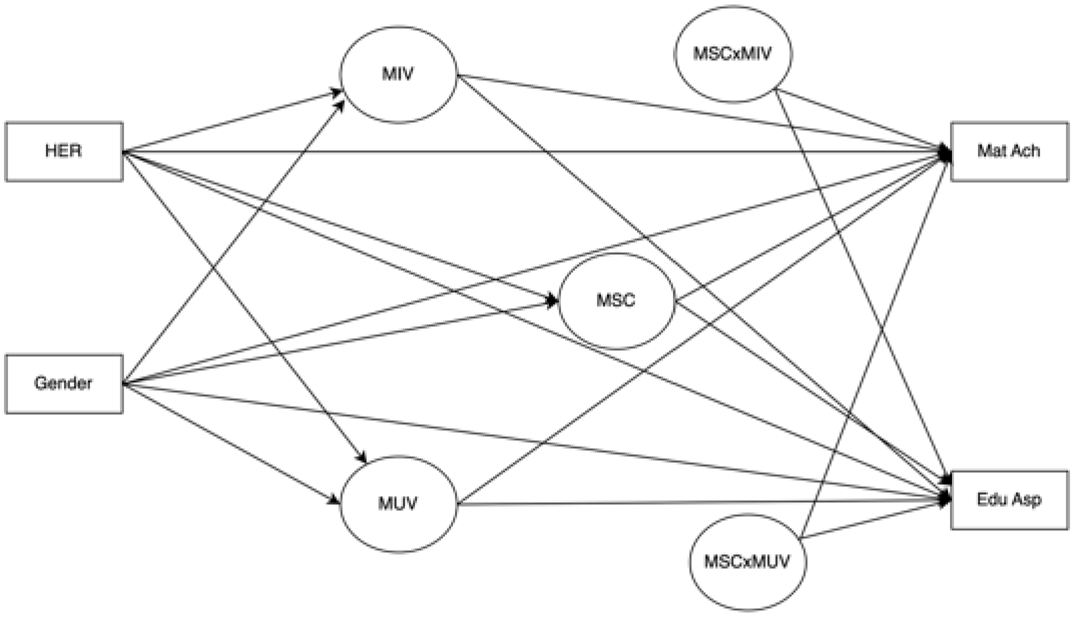
In this chapter, the moderated meditational model proposed in Figure 6.1 is tested using an SEM. To do this, the structural equation model with multiple mediators where measurement models for latent constructs are incorporated, including both the direct effects between the exogenous manifest variable HER, gender and endogenous latent variable mathematics achievement and educational aspirations (outcome), as well as the indirect effects from HER and gender via the latent mediators MSC, MIV, and MUV on the outcome variables mathematics achievement and educational aspirations. To be more specific for model specification in MPlus, mathematics achievement and educational aspirations are regressed on MSC, MIV, MUV, MSCxMIV, MSCxMUV, HER, and gender (direct effect), and MSC, MIV, and MUV are regressed on HER and gender (direct effect), and mathematics achievement and educational aspirations regressed on HER and gender via MSC, MIV, and MUV (specific indirect effect).

In the models employed in this study, there exists a correlation between the dependent and independent variables. Notably, the dependent variables, namely mathematics achievement and educational aspiration, were jointly estimated within the structural models, allowing for the consideration of their correlation. However, the correlation links between variables were not explicitly displayed in the figures depicting the models. This decision was primarily driven by the research's specific focus on regression analysis rather than emphasizing the correlation relationships.

By not showing correlation links in the figures, I aimed to maintain visual simplicity and enhance the ease of comprehension for readers. Displaying correlation relationships on the figures could potentially lead to visual complexity, making it more challenging for readers to grasp the critical insights from the models.

It is essential to emphasize that even though the correlation relationships are not visually presented in the figures, they were taken into account during the analysis. The existence of a correlation between dependent and independent variables was appropriately considered and accounted for in the statistical modelling and regression analysis conducted in the study. As such, the research findings and conclusions take into consideration the correlation relationships among the variables, despite not being explicitly depicted in the figures for the reasons explained above.

Figure 6. 1 The hypothetical conceptual framework of the mediating effect of EVT in the relationship between background factors and mathematics achievement and educational aspirations.



Note. MIV = mathematics intrinsic value; MSC = mathematics self-concept; MUV = mathematics utility value; MAT ACH = mathematics achievement; EDU ASP = educational aspirations; MSCxMIV = interaction variable of MSC and MIV; MSCxMUV = interaction variable of MSC and MUV.

6.4. Results

Model 16 was developed for the fourth and fifth research questions addressed in this chapter. This model investigated the complex relationship of EVT constructs with demographic constructs and educational outcomes.

This section starts with an assessment of the model fit and then concludes with the results of Model 16 for Japan, Türkiye, and England separately.

6.4.1. Goodness of model fit

In the hypothesised model (Figure 6.1), the effects of gender and HER (a measure of SES) on mathematics achievement and educational aspirations are mediated by expectancy and values (MSC, MIV, and MUV), and the latent interactions (MIVxMSC and MUVxMSC) influence the mathematics achievement and educational aspirations. The SEM model fitted the data partly well in our sample ($\chi 2$ =17,390.483, df = 2,187, CFI = 0.900, TLI = 0.885, RMSEA = 0.042). It is acceptable for CFI and RMSEA values to be within the acceptable range but for TLI values to be below the acceptable range (0.90). It is also necessary to remind ourselves that although there is no consensus, it is recommended that the RMSEA for model fit indicators should not be higher than 0.10, preferably less than 0.08, and CFI and TLI values higher than 0.90 (Hu and Bentler, 1999; Kline, 2015; Marsh et al., 2004). In line with Trautwein et al. (2012), I consider a model fit acceptable when at least two fit indices meet the typical range. Moreover, I also note that unique characteristics of data structures influence fit indices (e.g. Heene, Hilbert, Draxler, Ziegler, and Buhner, 2011), which invalidates sweeping generalisations regarding cut-off values (Trautwein et al., 2012).

The total amount of variance explained is 42%, 43%, and 35% for maths achievement and 17%, 21%, and 18% for educational aspirations in Japan, Türkiye, and England, respectively. There is less explained variance in educational aspirations than in mathematics achievement in all three countries. This may be due to the fact that the EVT constructs include specific items related to mathematics to measure students' attitudes towards mathematics whereas educational aspiration is a variable that indicates students' general future educational goals.

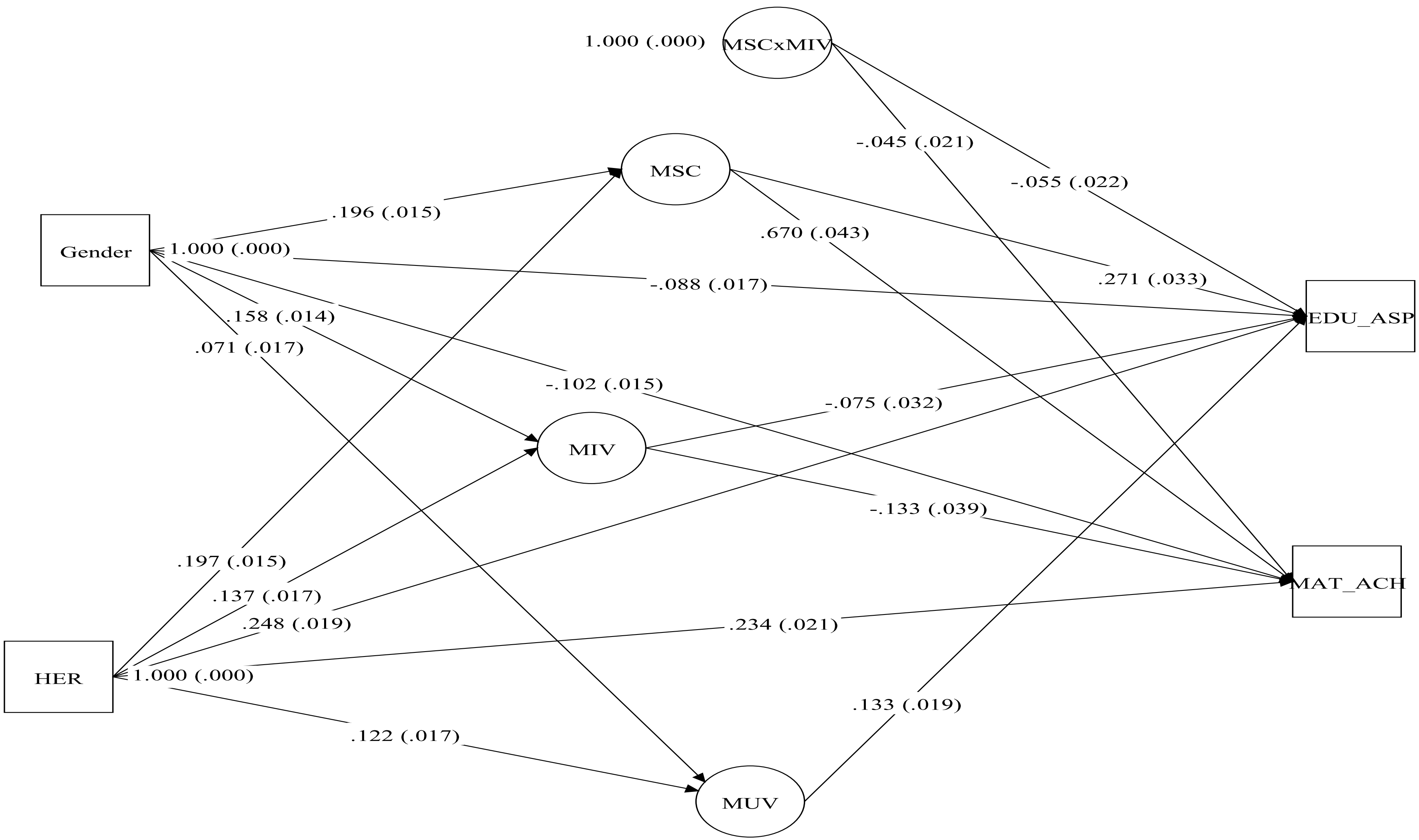
6.4.2. The relationship between gender, home educational resources (HER), EVT, and educational outcomes

The section is presented based on country results of the multigroup SEM analysis. In line with this, tables 6.1, 6.2 and 6.3 provide results for Japan, Türkiye, and England, respectively. In addition, the relationship of variables is demonstrated through diagrams for each country with significant paths in Figure 6.2 for Japan, Figure 6.3 for Türkiye, and Figure 6.4 for England. The figures demonstrate the effect sizes of the direct path coefficients of the standardised solutions, while the tables present the indirect path coefficient effect sizes.

Japan

Figure 6.2 shows that Japanese male students tend to have high expectancy-value beliefs over their female peers (MSC: $\beta = 0.196$; MIV: $\beta = 0.158$; MUV: $\beta = 0.071$). In contrast, the direct relationship between gender and mathematics achievement favours female students ($\beta = -0.102$). Interestingly, as shown in Table 6.1, the corresponding indirect path largely offsets the direct path from gender to mathematics achievement. These results suggest that boys are more likely to have higher EVT beliefs, which are associated with higher mathematics achievement (the indirect path from gender), whereas girls tend to perform better in mathematics when their EVT beliefs are similar to those of boys (the direct path from gender). Overall, however, there is no gender difference in mathematics achievement with respect to the total effect ($\beta = 0.011$).

Figure 6. 2 Structural model of Japan



Note. Standardised effect size (standard errors) for statistically significant paths was presented in the model for clarity. Note. MSC=mathematics self-concept; MIV= mathematics intrinsic value; MUV = mathematics utility value; HER = home educational resources; MAT_ACH = mathematics achievement; EDU_ASP = educational aspiration. MSC ×MIV = mathematics self-concept by intrinsic value interaction. 1= female, 2 = male

Table 6. 1 The direct, indirect, and total effects of gender and HER via EVT factors on educational outcomes in Japan

Outcome Variables	Direct effect of EVT	Indirect effect of EVT Via MSC	Via MIV	Via MUV	Total Indirect of EVT	Total effect
Gender						
Mathematics Achievement	-.102 (.015)	.131 (.013)	-.021 (.006)	.002[a] (.001)	.112 (.010)	.011[a] (.016)
Educational Aspirations	-.088 (.017)	.053 (.008)	-.012 (.005)	.010 (.003)	.051 (.006)	-.038 (.017)
Home Educational Resources (HER)						
Mathematics Achievement	.234 (.021)	.132 (.014)	-.018 (.006)	.004[a] (.002)	.118 (.011)	.352 (.019)
Educational Aspirations	.248 (.019)	.053 (.008)	-.010 (.004)	.016 (.003)	.060 (.006)	.308 (.018)

Note: MSC = mathematics self-concept; MIV = mathematics intrinsic value; MUV = mathematics utility value. 1 = female 2 = male. [a] represents an insignificant association. Standardised effect size (standard errors) is presented in the table.

As for educational aspirations, the direct path leading to girls having higher aspirations (β = -0.088) is only slightly countered by the corresponding indirect path leading to boys having higher educational aspirations. In other words, when girls and boys have similar MSC, girls have higher educational aspirations (β = -0.088), but since boys have higher MSC (β= 0.053), it reduces the gender gap in educational aspirations. Despite this, there is a small advantage in educational aspirations for girls in the total effect (β = -0.038).

As shown in Figure 6.2, there is evidence that HER has a positive direct effect on students' motivational beliefs and educational outcomes. This indicates that students with better HER are likely to be more motivated, realise higher mathematics achievement, and have higher educational aspirations. In particular, high HER is a statistically significant predictor of mathematics achievement (β = 0.234). In addition, as shown in Table 6.1, there is a significant and positive indirect path from HER to mathematics achievement, indicating that MSC has a positive role in mediating mathematics achievement (β = 0.132). It can therefore be concluded that an increase in HER promotes children's MSC and that this positive relationship results in increased mathematics achievement. Most likely, due to the high correlation between MSC and MIV, which we have previously discussed (see Chapter 5), the mediating effect of MIV has a detrimental effect on educational outcomes.

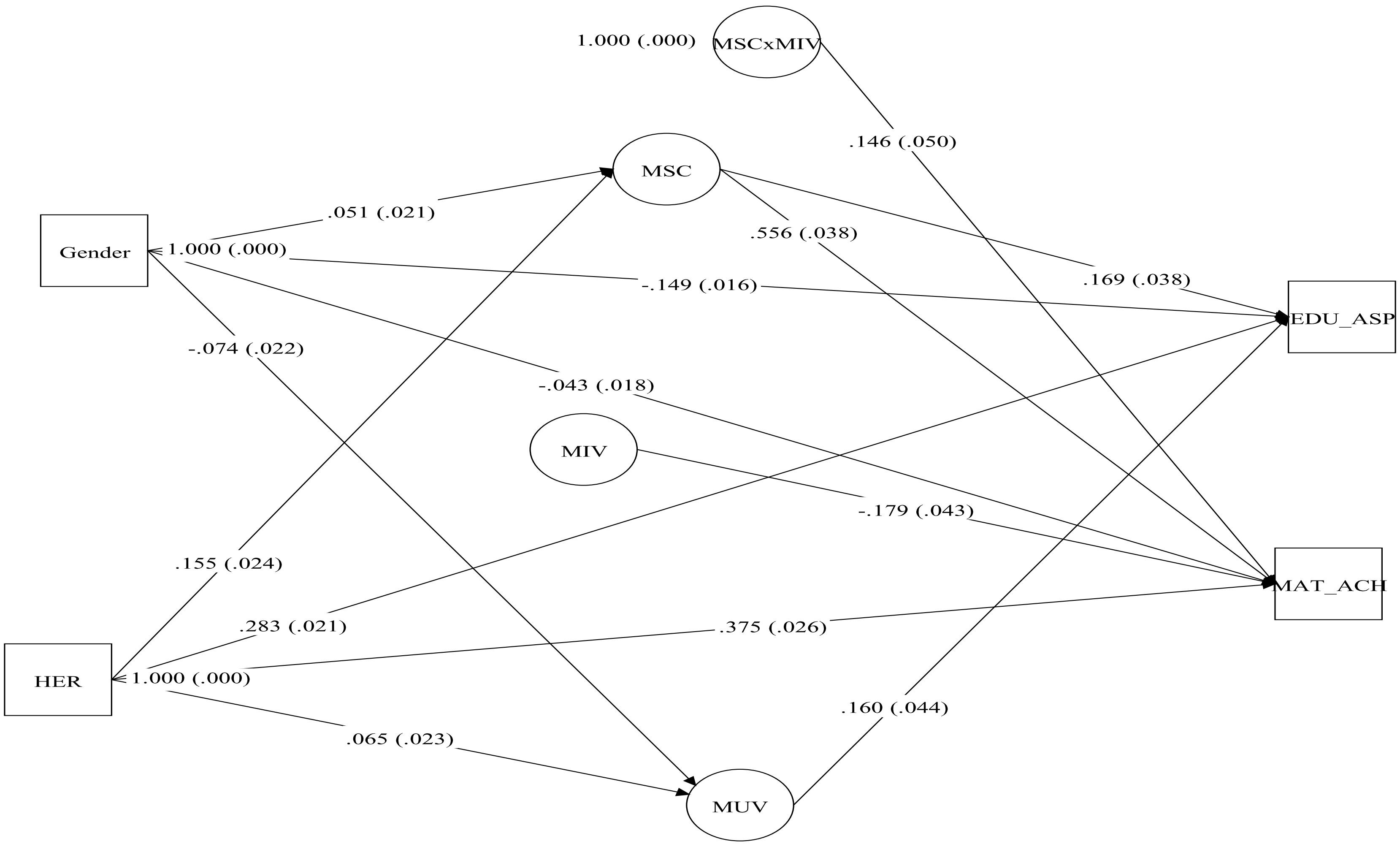
The relationship between HER and educational aspirations yields a similar result. Specifically, students with higher HER have higher educational aspirations (β =0.248). The relationship between HER and educational aspirations is positively mediated by MSC (β = 0.053), even if it is not as strong in mathematics achievement. On the other hand, when the direct effect of HER on mathematics achievement and educational aspiration is considered, the impact of HER on educational aspiration (β = 0.248) is slightly higher than that of mathematics achievement (β = 0.234). At this point, EVT factors seem to mediate more strongly between HER and achievement. However, HER is a stronger predictor of mathematics achievement (β = 0.352) than educational aspirations (β = 0.308) in terms of the total effect. As mentioned earlier, to possible reason for this is that EVT factors include items specific to mathematics, whereas educational aspiration is about general educational goals.

In the relationship between motivation factors and achievement, MSC is the strongest predictor among motivation factors (β = 0.670). Contrary to expectations, MIV has a negative effect on achievement (β = -0.133). As underlined in Chapter 5, the most likely reason for this negative effect is the high correlation between MIV and MSC. In addition, while MUV has no significant effect on achievement, it has a significant effect on educational aspiration (β = 0.133). Moreover, the MSCxMIV interaction also has a statistically significant effect on achievement (β = -0.45) and educational aspirations (β = -0.55). The relationship between EVT factors and educational outcomes, as well as the interaction effect, is explained in detail in Chapter 5.

Türkiye

The findings in Figure 6.3 indicate that gender influences motivational beliefs directly, with boys showing higher MSC (β = 0.051) and girls higher MIV (β = -0.074), but there are no significant differences in MUV between boys and girls. On the other hand, as indicated in Table 6.2, the MSC has a mediating effect (β = 0.030) in favour of male students in the relationship between gender and mathematics achievement, which means an indirect path offsets the path from gender to mathematics achievement. Based on these findings, it is likely that boys could have higher MSC beliefs, which are indirectly associated with higher mathematics achievement.

Figure 6. 3 Structural model of Türkiye



Note. Standardised effect size (standard errors) for statistically significant paths is presented in the model for clarity. Note. MSC = mathematics self-concept; MIV = mathematics intrinsic value; MUV = mathematics utility value; HER = home educational resources; MAT_ACH = mathematics achievement; EDU_ASP = educational aspiration. MSC ×MIV = mathematics self-concept by intrinsic value interaction. Gender: 1 = female, 2 = male

Table 6. 2 The direct, indirect, and total effects of gender and HER via EVT factors on educational outcomes in Türkiye

Outcome Variables	**Direct effect of EVT**	**Indirect effect of EVT**			**Total indirect effect of EVT**	**Total effect**
		Via MSC	**Via MIV**	**Via MUV**		
Gender						
Mathematics Achievement	-.043 (.018)	.030 (.012)	.002[a] (.004)	-.004[a] (.003)	.026 (.010)	-.017[a] (.021)
Educational Aspirations	-.149 (.016)	.009 (.004)	.001[a] (.001)	-.012 (.005)	-.003[a] (.007)	-.152 (.017)
Home Educational Resources (HER)						
Mathematics Achievement	.375 (.026)	.086 (.015)	.001[a] (.004)	.004[a] (.003)	.091 (.013)	.467 (.029)
Educational Aspirations	.283 (.021)	.026 (.008)	.001[a] (.001)	.010 (.004)	.037 (.010)	.320 (.022)

Note: MSC= mathematics self-concept; MIV = mathematics intrinsic value; MUV = mathematics utility value. 1= female, 2 = male. [a] represents an insignificant association. Standardised effect size (standard errors) is presented in the table.

On the other hand, when female and male students' MSC beliefs are similar, girls may have higher mathematics achievements than males. Female students outperformed male students in mathematics achievement with a statistically small effect size in direct effect (β = -0.043), but the mediating effect of MSC (β = 0.030) in favour of male students reduced the gap in mathematics achievement between male and female students. This implies that there is no gender difference in terms of the total effect, but there is an indirect effect of MSC that favours boys.

Regarding educational aspirations, it can be seen in Table 6.2 that the direct path from gender to educational aspirations favours female students (β = -0.149). Although MSC has a very small but significant indirect effect in favour of boys, the total effects show results in favour of girls (β = -0.152). It is also seen that female students have more MUV than male students, albeit with a small effect (β = -0.012).

It is evident from Figure 6.3 that the positive direct effects of HER on student motivation, mathematics achievement, and educational aspirations indicate that students who live in a high HER environment are likely to have more positive motivation, higher mathematics achievement, and higher educational aspirations. Contrary to expectations, however, this study found that HER has no significant direct effect on MIV. The indirect path from HER to mathematics achievement has a significant effect via the mediation of MSC ($\beta = 0.086$). This result suggests that high HER has both a substantial direct effect on mathematics achievement and an indirect effect on mathematics achievement by enhancing the student's MSC beliefs.

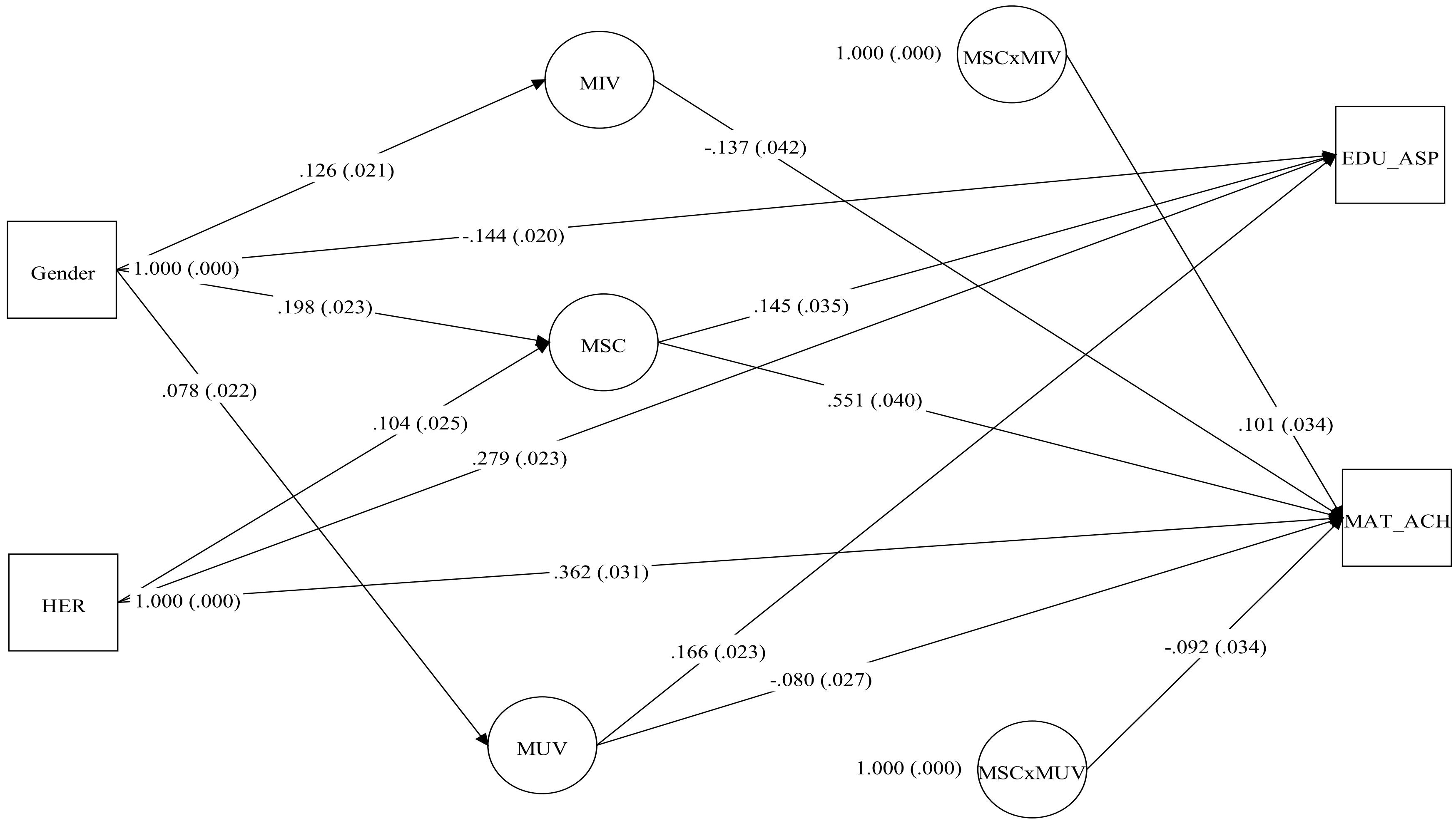
The effect of HER on educational aspirations is mediated via both MSC and MUV. As Table 6.2 illustrates, the regression coefficient between HER and educational aspirations ($\beta = 0.283$), that between MSC and educational aspirations ($\beta = 0.026$), and that between MUV and educational aspirations ($\beta = 0.010$) are significant. This result is consistent with our expectations (H11), indicating that HER is an important factor in predicting educational aspirations directly and indirectly by promoting students' MSC and MUV.

In the relationship between motivation factors and achievement, MSC is the strongest predictor among motivation factors ($\beta = 0.556$). Contrary to expectations, MIV has a negative effect on achievement ($\beta = -0.179$). As underlined in Chapter 5, the most likely reason for this negative effect is the high correlation between MIV and MSC. In addition, while MUV has no significant effect on achievement, it has a significant effect on educational aspiration ($\beta = 0.160$). Besides, the MSCxMIV interaction also has a statistically significant effect on achievement ($\beta = 0.146$). The relationship between EVT factors and educational outcomes, as well as the interaction effect, is explained in detail in Chapter 5.

England

Figure 6.4 shows that there is no difference in maths achievement between boys and girls. However, the direct effect of gender on expectancy-value beliefs reveals that boys typically seem to have a high MSC (β = 0.198), MIV (β = 0.126), and MUV (β =0.078). It appears that boys are more likely to possess higher MSC beliefs and therefore are more likely to perform better in mathematics (the indirect path from gender), whereas it appears that girls and boys perform equally well in mathematics when their MSC beliefs are similar (the direct path from gender). Specifically, a strong mediation effect of MSC in favour of male students emerged (β = 0.109). In summary, the effects of the relationship between gender and mathematics achievement are indirect, which means that, even though there are no direct effects of gender on mathematics achievement, the relationship between gender and mathematics achievement is mediated by MSC (see Table 6.3).

Figure 6. 4 Structural model of England



Note. Standardised effect size (standard errors) for statistically significant paths is presented in the model for clarity. Note. MSC = mathematics self-concept; MIV = mathematics intrinsic value; MUV = mathematics utility value; HER = home educational resources; MAT_ACH = mathematics achievement; EDU_ASP = educational aspiration; MSC ×MIV = mathematics self-concept by intrinsic value interaction; MSCxMUV = mathematics self-concept by utility value interaction. 1= female, 2 = male.

Table 6. 3 The direct, indirect, and total effects of gender and HER via EVT factors on educational outcomes in England

Outcome Variables	Direct effect of EVT	Indirect effect of EVT			Total indirect effect of EVT	Total effect
		Via MSC	Via MIV	Via MUV		
Gender						
Mathematics Achievement	-.036[a] (.029)	.109 (.016)	-.017 (.006)	-.006 (.003)	.085 (.013)	.049[a] (.033)
Educational Aspirations	-.144 (.020)	.029 (.008)	.002[a] (.004)	.013 (.002)	.043 (.008)	-.101 (.021)
Home Educational Resources (HER)						
Mathematics Achievement	.362 (.031)	.057 (.015)	-.007[a] (.004)	-.003[a] (.002)	.047 (.012)	.409 (.033)
Educational Aspirations	.279 (.023)	.015 (.006)	.001[a] (.002)	.007[a] (.004)	.023 (.008)	.302 (.023)

Note: MSC= mathematics self-concept; MIV = mathematics intrinsic value; MUV = mathematics utility value. 1 = female, 2 = male. [a] represents an insignificant association. Standardised effect size (standard errors) is presented in the table.

With regard to educational aspirations, it can be noted that the direct path from gender to educational aspirations favours female students (β = -0.144). Even though MSC (β = 0.029) and MUV (β = 0.013) have significant indirect effects in favour of boys, they also have significant direct effects in favour of girls both directly (β = -0.144) and totally (β = -0.101).

As shown in Figure 6.4, there is evidence that HER has a positive direct effect on mathematics self-concept (β = 0.104), mathematics achievement (β = 0.362), and educational aspirations (β = 0.279). This finding suggests that students with high levels of HER are likely to have a higher self-concept, mathematics achievement, and educational aspirations than those with low levels of HER. Furthermore, Table 6.3 shows that in relation to HER, mathematics achievement and educational aspirations are mediated by MSC, suggesting that HERs influence MSC and promote mathematics achievement and educational aspirations.

In the relationship between motivation factors and achievement, MSC is the strongest predictor among motivation factors (β = 0.551). Contrary to expectations, MIV has a negative effect on achievement (β = -0.137). As underlined in Chapter 5, the most likely reason for this negative effect is the high correlation between MIV and MSC. In addition, while MUV has a negative significant effect on achievement (β = -0.080), it has a positive significant effect on educational aspiration (β = 0.166). Furthermore, the MSCxMIV and MSCxMUV interaction also has a statistically significant effect on achievement. The relationship between EVT factors and educational outcomes, as well as the interaction effect, is explained in detail in Chapter 5.

6.5. Summary

This chapter aimed to examine how the relationship between HER, gender, and educational outcomes (mathematics achievement and educational aspirations) is mediated by the expectancy-value components, which are mathematics self-concept, and intrinsic and extrinsic values.

The results are generally in accordance with our expectations, although some are contrary to those expectations. Our results show that gender did not have a significant effect on mathematics achievement in all three countries. Our study reveals, however, that female students have higher educational aspirations than their male counterparts in all three countries. As expected, these findings are consistent with the recent gender difference observed in educational aspirations favouring girls, which is in line with our expectations (Guo et. al, 2015; Schoon and Polek, 2011). It is also important to note that male students have higher MSC, MIV, and MUV beliefs in England and

Japan than female students. Albeit only a slight positive difference in favour of male students in MSC beliefs, females have a small advantage in MUV beliefs in Türkiye. Furthermore, there is no gender difference in MIV for Türkiye. Therefore, we can conclude that while male students are more motivated than female students in England and Japan to learn mathematics, a similar gender gap cannot be found in Türkiye.

In parallel with our expectations, HER contribute directly and indirectly to mathematics achievement and educational aspirations. Another noteworthy point is that the mathematic self-concept is the most powerful mediator of mathematics achievement and educational aspirations among the EVT components. In addition, it plays an important role in promoting boys' achievement in mathematics. Contrary to our expectations, the findings indicate that there is no relationship between HER and MIV in either Türkiye or England.

In the next chapter, all the results from Chapters 4, 5, and 6 will be discussed in the context of the wider literature and theory in order to answer the research questions in full.

Chapter 7 Discussion

The discussion chapter is organised according to the research questions (see Section 2.4 in Chapter 2). Accordingly, each research question and its subparts are discussed in light of my findings, and relevant literature and theory. The first two research questions analyse the psychometric properties of the TIMSS 2019 motivational items and compare measurement models based on CFA and ESEM (Chapter 4). The third research question investigates the predictive power of EVT factors for mathematics achievement and educational aspiration and the interaction effect of expectancy and value constructs on these dependent variables (Chapter 5). Finally, the last two research questions, RQ4 and RQ5, investigate the mediating role of EVT constructs in the relationship between demographic variables (gender and HER) and educational outcomes (mathematics achievement and educational aspirations) (Chapter 6).

7.1. What are the psychometric properties of TIMSS 2019 motivation measures?

In this research question and corresponding hypotheses, reliability, the goodness of model fit and negative item effect, factor structure, CFA and ESEM comparison, and measurement invariance are investigated. I take each of these in turn.

7.1.1. Reliability

In the reliability analysis, McDonald's omega was calculated in addition to Cronbach's alpha value provided in the TIMSS 2019 technical report. Some studies in the literature have found that TIMSS motivation measures are unreliable in non-Western countries (Bofah and Hannula, 2015; Marsh et al., 2013a; Rutkowski and Rutkowski, 2010). Therefore, in this context, I hypothesised that the TIMSS 2019 motivational constructs would meet the reliability criteria, but that England would present a higher reliable index than Japan and Türkiye. For example, Marsh et al. (2013a) found that the reliability of mathematics motivation constructs was higher in Western countries than Arab countries. However, the results of this study, contrary to the literature, provide strong evidence that TIMSS motivation measures are reliable measures based on alpha and omega estimation in Türkiye, England, and Japan (see Section 4.3.2 in Chapter 4). Additionally, the omega value is estimated in this study by considering the effect of negatively worded items. The reliability indices may overestimate results when this issue is not considered (Bofah and Hannula, 2015). In this study, alpha and omega indices, where the negative item effect was not considered, are very close to each other and provided high reliability values. However, when the omega value was calculated by correlating the error terms of negative items with three items in MSC and two items in MIV, thus controlling the negative item effect, it provided more realistic reliability scores. In line with the literature, this study concludes that under certain situations, such as the analysis of latent constructs with negative items, the alpha may overestimate the reliability and might not provide an appropriate measure (Bentler, 2009; Fu et al., 2022; Raykov, 2001; Yang and Green, 2011). Therefore, the omega reliability test may be preferred in such cases.

7.1.2. The goodness of model fit and method effect: negative item effect

This study found a substantial method effect associated with negatively worded items to measure motivational constructs in TIMSS 2019 (Section 4.3.3.1 in Chapter 4). In accordance with the literature and our hypothesis, the findings of the model fit and reliability estimation confirm the claim that negatively worded items can adversely affect the validity and reliability of scales (Bofah and Hannula, 2015; Marsh et al., 2013a; Raykov, 2001). It was determined that measurement models that did not consider the effects of negatively worded items fit the data once they established a correlation between error terms of negative items within the measurement model. In other words, when negative item effects are not controlled by correlating their error terms, the measurement model is poorly defined and does not meet the model fit criteria (see the comparison of model fit with and without negative item effects in Table 4.3 in Chapter 4). Most of the secondary analyses conducted with TIMSS have used manifest scores (i.e. observed variables), but it not simple to incorporate these method effects into analyses based on manifest scores (Marsh et al., 2013a). Therefore, the result of this study agrees with the recommendation of Marsh et al.,(2013a), which encourages researchers to use latent variable models rather than methods based on manifest scores, to minimise biased measurement effects due to negative items.

The adverse effects of negatively worded items on model fit have long been known, particularly for children as well as for adults and adolescents (Hooper et al., 2013; Marsh, 1986). Moreover, negative item effects have been found in the literature to be significantly higher in countries with low achievement levels, possibly because students have poor reading comprehension and cannot fully understand negatively worded items (Bofah and Hannula, 2015; Hooper et al., 2013; Metsämuuronen, 2012). The possible reason for the method effect may be related to the extent to which reading, and comprehension proficiency is affected by the negatively worded items. There is, however, a method effect presents in not only underperforming countries but in all countries, although the effect differs by country. As a result of the study conducted by Michaelides (2019) with TIMSS motivation data, fourth-grade students with low reading ability responded differently to negative and positive items in such a way that their total scores deviated downwards. This is also in line with Marsh's (1986, 1996) findings that method effects are reduced for more verbally competent and older students. Hence, this result could be explained by difficulties processing negatively worded items. Thus, it may be regarded as an aspect of the cognitive development phenomenon. This study demonstrates a negative item effect in all three countries possibly due to the age of eighth-grade students, as expected (Marsh, 1996). Nevertheless, in accordance with the literature, the reliability index calculated with and without the negative item effect (see Table 4.2 in Chapter 4) indicates that the highest effect can be found in Türkiye, where achievement is the lowest among the countries examined in this study (Bofah and Hannula, 2015; Hooper et al., 2013; Metsämuuronen, 2012). Overall, our findings suggest that method effects, if not handled explicitly, are not only likely to obscure the underlying structure of these scales but also cause a bias through an unreliable estimation process and poor model fit.

7.1.3. Factor structure, CFA and ESEM comparison, and measurement invariance

In TIMSS 2019, students' motivations for mathematics were measured in the context of student attitudes towards mathematics through three latent constructs: student confidence in mathematics, students like learning mathematics, and student value of mathematics (Mullis et al., 2020). Therefore, I hypothesised that the factor structure of motivation measures supports the a priori structure (derived from TIMSS 2019 motivation scale) that they intend to measure. After negative item effects were considered, the CFA and ESEM models were developed to test this, and both models clearly support the a priori factor structure. The ESEM model provided better model fit values than CFA.

There could be many cross-loadings in a measurement instrument (although they are less significant than main loadings) that are consistent with the underlying theory. Otherwise, by setting cross-loading to zero, researchers might specify a parsimonious model that does not adequately fit the data when using the independent cluster model of CFA (Guay et al., 2015). More importantly, when cross-loadings, even small ones, are not estimated, then the only way to represent these associations between specific indicators and other constructs is through the latent factor correlations, which end up being overestimated in many applications of CFA (e.g. Asparouhov and Muthén, 2009; Marsh et al., 2009; Morin et al., 2013). As discussed in the literature, ESEM overcomes the limitations of CFA by estimating all cross-loadings between indicators and latent constructs. In this study, I only allowed cross-loading between intrinsic and utility value measures since self-concept is considered a separate construct theoretically (Marsh et al., 2010). It was consistent with theoretical expectations to find cross-loadings between intrinsic and utility value indicators using the ESEM approach. Further, it is evident from the literature that ESEM structures are more closely aligned with the theoretical conceptualisation of factors (Alamer, 2021; Gomez et al., 2020; Guay et al., 2015; Marsh et al., 2020; Marsh et al., 2010; Marsh, Herbert W. et al., 2014; Morin et al., 2013; van Zyl and Ten Klooster, 2021; Xiao et al., 2019). For instance, the theoretical framework of this study, EVT, considered intrinsic and extrinsic value as a subcomponent of task value (Eccles and Wigfield, 1983). Even though intrinsic and extrinsic value can be distinguished in theory, in most applications, it is not as black and white as it appears in theory (Guay et al., 2015). The results are consistent with recent findings and our hypothesis, indicating that ESEM tends to provide a better model fit factor when cross-loadings are present in the population model but remains unbiased when cross-loadings are absent (Asparouhov et al., 2015).

Since the ESEM model provided better results than the corresponding CFA, the analyses after this stage were performed with ESEM. In the measurement invariance test conducted within the ESEM framework, though there was strong evidence for configural and metric invariance, partial evidence for scalar invariance was found among the three countries. Scalar invariance is important for comparing latent means (Marsh et al., 2013a). In the scalar invariance test, the intercepts of some items differed across countries, so these items were freely estimated to achieve partial scalar invariance. When these items were examined, it could be seen that there were negatively coded items in the MSC and MIV measures. Therefore, as in the model fit and reliability test, the negative items are likely to cause bias in scalar invariance. The findings of this study agree with Chiu's (2012) argument that negatively worded items can be unreliable in cross-cultural studies.

7.2. What is the correlational relationship between motivation factors, gender, HER, and educational outcomes (mathematics achievement and educational aspirations)?

In this research question, the latent mean comparison across countries and the relationship between EVT constructs and demographic indicators and educational outcomes are investigated.

7.2.1. Latent mean comparison

The results of this study show that the mean value of motivational beliefs is highest in Türkiye and lowest in Japan. However, in mathematics achievement, the opposite is the case, and Japan has the highest mean achievement score, while Türkiye has the lowest (see Chapter 4). In this context, the results of this study seem "perplexing" and "paradoxical", but they are in line with other large-scale survey studies (Bofah and Hannula, 2015; Marsh et al., 2013a; Shen and Tam, 2008). There was a positive correlation between students' motivational constructs (such as self-concept) and their achievement at the individual level but a negative correlation at the country level (Ker, 2017). The frame of reference effect can partly explain this “paradoxical” and “perplexing” situation based on self-concept theory (Marsh, 2007). Clearly, frame of reference effects significantly influenced motivational constructs (Marsh, 2007; Marsh et al., 2008). Thus, Turkish students develop their self-concepts in relation to other Turkish students rather than Japanese, English, or students from other countries. Hence, achievement in Türkiye is similarly related to motivation constructs as elsewhere; neither are motivational constructs lower than in countries with high performance. On the contrary, Japanese students are not likely to compare themselves with students from other countries but rather with their classmates and schoolmates around them. As a result of comparing their mathematics achievements with their peers who have relatively higher maths achievements on average, they may see themselves needing improvement. Shen and Pedulla (2000, p.237) offered a related suggestion by arguing that this pattern may reflect “low academic expectations and standards in low-performing countries and high academic expectations and standards in high-performing countries” in more comprehensive terms. Parallel to this, Marsh et al., (2006) suggested that cultural differences affect the way in which one expresses positive things about oneself, such as self-concept. A similar phenomenon has been suggested by Minkov (2008) in relation to the cultural value of “monumentalism” in terms of self-enhancement versus self-effacement, and self-stability/consistency versus self-flexibility, as well as the need to reinforce self-improvement through high performance. The findings of these studies suggest that monumentality at the national level is positively associated with positive self-beliefs but negatively associated with achievements.

7.2.2. The relationship among motivational constructs, background variables, and educational outcomes

The discriminant validity of measures with multidimensional constructs plays a critical role in their construct validity and usefulness (Marsh et al., 2013a). More attention needs to be paid regarding the distinction between self-concept and intrinsic value in applied self-concept and motivation research and in the theoretical models that underpin this applied research, particularly given the high correlation between these concepts. Accordingly, recent research has typically found a correlation between self-concept and intrinsic value of 0.7 or greater (Marsh et al., 2013a; Meyer et al., 2020; Nagengast et al., 2011). Despite the high correlations between MSC and MIV determined in this study, their discriminant validity is also supported by my work. In particular, consistent with a priori predictions, there is a substantial correlation between achievement and MSC, but the achievement and MIV correlation are weaker than MSC and achievement, especially for Türkiye and England. In addition, MSC has a higher correlation with educational aspiration than MIV (see section 4.3.4.2 in chapter 4).

Gender differences in achievement, educational aspirations, and motivation are particularly interesting in this study. This study supports the conclusions in the literature concerning student gender and mathematics achievement that there are no statistical differences between males and females in mathematics achievement in this age group (Else-Quest et al., 2010) in Japan and England but a small difference in favour of females in Türkiye (Hyde et al., 1990). In addition, in line with the literature, there is also a correlational difference between gender and educational aspiration in favour of female students in Türkiye and England, but not in Japan. Regarding the associations between gender and EVT constructs, Japanese and English male students have a higher correlation with mathematics motivation and MSCs than female students. On the other hand, in Türkiye, there is a more balanced result. While there is no difference between male and female students in terms of MIV, the MSC of male students and the MUV of female students are higher, concurring with other literature (Schoon and Polek, 2011; Guo et al., 2015).

Taken all together, although there is no association between gender and mathematics achievement in Japan and England, there is a correlation between gender and motivation in favour of boys. However, in Türkiye, there is a correlation with achievement in favour of girls, albeit slight, as well as a correlation between gender and motivational beliefs in favour of girls in MUV and in favour of boys in MSC. This can be explained by the gender equality paradox, which states that in countries with greater gender equality, there are differences in self-beliefs to the disadvantage of girls (Else-Quest et al., 2010; Guo et al., 2019; Marsh et al., 2021; Marsh et al., 2019). This may be due to the fact that female students in developed countries have more freedom in course selection regardless of economic concerns compared to their counterparts in less developed countries. Thus, in contrast, female students in economically less developed countries may feel pressured to pursue STEM fields by their families and internally since STEM is concerned the safer option for economic success (Else-Quest et al., 2010).

7.3. What is the relationship between student motivational factors and educational outcomes?

This research question examines the influence of expectancy (MSC) and task values (MIV and MUV) and the effect of their interaction on mathematics achievement and educational aspirations. Structural equation modelling (SEM) was applied for the analysis, and interaction terms were formed using an unconstrained approach guided by Marsh et al., (2004) to measure the interaction effect. The results are presented in Chapter 5. Four different SEM models were developed for the analyses, and the results are presented in Table 5.1 for EVT and mathematics achievement, and in Table 5.2 for educational aspiration in Chapter 5.

7.3.1. The additive effect of EVT

Based on the two models (M12 and M13, see Chapter 5), MSC appears to be more closely associated with achievement than value beliefs (MIV and MUV) (see Wigfield et al., 2009). However, the presence of value beliefs is an important predictor of other outcomes, such as educational aspirations (Guo et al., 2015; Nagengast et al., 2011). As MSC and MIV are closely associated, MIV loses predictive power when MSC is also included in the regression equation, which is a common issue according to the literature (see Guo et al., 2015; Lauermann et al., 2017; Meyer et al., 2019; Trautwein et al., 2012). Japan is the country with the highest variance ratio explained by the EVT structures. On the other hand, Türkiye and England have similar variance-explained ratios.

There is growing recognition that culture has had an impact on learning and motivation (Salili and Hoosain, 2007). As suggested by Hernandez and Iyengar (2001), people need the support of their cultures and beliefs to be motivated. Tonks et al. (2018) discussed the ways in which cultural background and culture can be included within the EVT model. It might be helpful to include various aspects of culture in the "Cultural Milieu" box; for example, "individualism" and "collectivism" may be considered broad cultural characteristics, while more specific interaction styles and processes among parents and children could be included. For instance, Ng (2003) argued that collectivist cultures and corresponding school practices shape Asian societies' perspectives on motivation. As a result of these views, parents and teachers in Asia often believe that learning is inextricably linked to achievement and that achievement is, at the same time, regarded as a social obligation. The author also argues that such cultural influences place a lesser emphasis on personal interests and enjoyment, compared with performance and achievement being significantly influenced by external factors. A similar argument was made by Markus & Kitayama (1991), who argued that in Asian societies, motivation tends to be derived from what benefits others or a group, whereas, in Western cultures, people tend to have self-interested motivations that benefit them directly. Regarding mathematics, Leung (2001) argued that educators in the West place high value on intrinsic motivation, but in contrast, extrinsic motivation (such as exam pressure) is considered a negative factor (such as exam pressure), whereas extrinsic factors (such as pressure) are more tolerated in the East Asian countries. In a way, this is one of the factors explaining why East Asian students report a lower level of motivation despite performing academically better than their counterparts in international comparison studies (PISA and TIMSS). The results of this study support these arguments by showing that utility value is significantly positive in Japan, a typical Asian country, but statistically insignificant in England, a typical Western country. These results also align with the literature as Zhu and Leung (2011) found that MIV has a positive effect on achievement, but MUV has a negative effect on achievement in Western countries.

In addition, Tonks et al. (2018) discuss how a given construct (e.g. MSC) may have different meanings for different cultures, indicating that the relationships specified in the model may differ substantially. Therefore, in order to understand how academic motivation functions and affects students from different cultures, it is essential to consider the effects of cultural values, norms, and practices. In the absence of such consideration, we may misguide students from different cultures and motivate them in a culturally inappropriate way. As Leung (2001) points out, such a difference in perception might be a result of the different views on human nature held by people from the East and the West. A proper level of pressure could well assist in directing students' attention and energy towards studying, since East Asians tend to believe that humans need some form of "push" in order to learn (Leung, 2001). In contrast, in the West, they are more likely to believe it is more important to stimulate students' interest to start learning.

The effect of utility value on this level of mathematics achievement in Türkiye should be investigated further. The possible reason for this situation may be that STEM fields offer a more prosperous life in economic and social terms and accordingly bring family and internal pressure. For example, a study by Mullis et al. (2016) found that 82% of Turkish parents exhibited positive attitudes toward mathematics and science among the participants in TIMSS 2015.

On the other hand, in line with the literature (Guo et al., 2015; Nagengast et al., 2011; Nagengast et al., 2011; Meyer et al., 2019; Trautwein et al., 2012), MSC was found to have the highest association with mathematics achievement among the EVT factors for all three countries in this study (for more discussion see section 7.2.2).

7.3.2. The effect of expectancy and task values interactions

One of the key contributions of this study is to test the interaction of expectancy and task value in the EVT theoretical framework. In this context, the interaction effects of MSC and task value were tested in this study. This study proved that the interaction of MSC and task value has a statistically significant effect on mathematics achievement, even at different levels and at different levels of instruction in the three countries (see for more detail section 5.4 in chapter 5). This is important because modern EVT (Eccles et al., 1983) usually ignores the expectancy task value interaction that was present in the original EVT (Atkinson, 1957). However, most of the studies conducted within the framework of modern EVT are non-experimental – see Trautwein et al. (2012) who state that non-experimental studies have two main disadvantages in terms of identifying interaction effects; namely, (1) scores for predictor variables are generally normally distributed, and (2) there is measurement error in analyses based on manifest scores in the measurement of predictor and criterion variables in non-experimental studies. This means that there is typically insufficient statistical power to detect interaction effects unless these effects are unusually large. However, in recent years, international large-scale assessments programmes such as TIMSS and PISA have provided large data for researchers, and the development of latent models and interaction analysis approaches within the framework of SEM where measurement error is minimal, such as latent moderated SEM and unconstrained product indicator approach have eliminated these limitations.

The interaction of MSC and MIV is analysed in model 14 and illustrated in figure 5.2. Figure 5.2 simply revealed that a downward trend is evident in mathematics achievement among students in England and Türkiye with low MSCs and high MIVs. Furthermore, a high MIV has little effect on achievement for students with high MSC in England and Türkiye. Therefore, a low MSC will not compensate for a high MIV but will adversely affect the student's mathematics achievement. In Japan, students with a high MSC level contribute more to mathematics achievement when their MIV is low. Thus, higher levels of intrinsic motivation in students with high mathematics self-concept negatively affect mathematics achievement in Japan. When all countries are considered comparatively, MIV has a detrimental effect as the MSC level decreases in Türkiye and England, while the opposite is the case in Japan, where the increase in MIV has a negative effect as the MSC level rises. In other words, the negative effect of MIV is higher for Turkish and English students with low MSC, while it is more for students with high MSC in Japan. The negative effect of MIV on the interaction with MSC is also consistent with the recent literature (Meyer et al., 2020; Trautwein et al., 2012). On the other hand, a similar picture emerges in the impact of the MSCxMIV interaction on the educational aspirations of Türkiye and England (see figure 5.4). The only difference that emerges here is that for students with high MSC, high MIV has a positive effect, not a negative in contrary to mathematics achievement. However, in other cases, the increase in MIV in parallel with the decrease in MSC causes a negative effect on educational aspirations. In Japan, for students with a high MSC, a higher MIV has a negative impact on educational aspirations.

The interaction effect of MSC and MUV was investigated in Model 15, and the relationship with mathematics achievement in figure 5.3 and the relationship with educational aspirations in figure 5.5 was visualised. MSC and MUV exhibited a significant interaction effect in Japan and Türkiye, whereas the effect was insignificant in England for mathematics achievement (see figure 5.3). An overview of figure 5.3 shows that students with higher MSCs tend to do better in mathematics regardless of the level of MUV for Japan and Türkiye. Specifically, however, for Japan, the increase in the MUV of students with low MSC is positively reflected in mathematics achievement, while it is almost ineffective for students with high MSC. For Türkiye, the opposite is the case, the increase in MUV is virtually insignificant for students with low MSC; however, the rise in MUV for students with high MSC promotes mathematics achievement. It was also found that the mathematics utility value and the mathematics self-concept had a significant interaction effect on educational aspiration only in Türkiye (see figure 5.5). Utility value enhances students' self-concept, which leads to a positive increase in educational aspirations.

In Chapter 5, we specifically report the results of the expectancy-value interaction effect by country and variable. It is evident from the findings that the interaction variable has only a very small effect on many models, and this is in line with most literature (Guo et al., 2015; Nagengast et al., 2013; Trautwein et al., 2012); however, what does this mean for a theoretical and practical interpretation of the interaction effect? As a point of emphasis, it should be noted that in many models, expectancy and task value interaction variables have a larger effect size than MUV (see section 5.4 in chapter 5). Furthermore, the interaction variable (MSCxMIV and MSCxMUV) contributes between 1% and 15% in the explained variance ratios. It should also be noted that there may have been a negative effect on the interaction variable due to the high correlation between the MSC and the MIV. Therefore, as Nagengast et al. (2013) pointed out, the results of this study should not be interpreted as indicating that interaction terms are of no theoretical significance to EVT. On the contrary, the presence of significant interactions indicates that the two predictor variables might be related in a multiplicative manner (Arnold and Evans, 1979; Busemeyer and Jones, 1983), regardless of predictive power. This study provides evidence for the interaction effect of expectancy and task value. It should be emphasised that the contribution of the interaction variable to mathematics achievement, which is produced from the multiplicative of these constructs (interaction), is more important than the statistical effect size. In particular, this study revealed the positive effect of increasing students' MSCs and MIVs on mathematics achievement. This study also justifies applying latent interaction modelling because models with manifest variables would likely yield even smaller effect sizes, potentially resulting in the unjustified rejection of multiplicative relations in error due to the small interaction effects (Meyer et al., 2020).

7.4. What is the relationship between gender mathematics achievement and educational aspirations? Do EVT factors mediate this relationship?

The prominent results of this research question are that gender is a statistically significant predictor of achievement in favour of female students in Japan and Türkiye, whereas gender has no effect on mathematics achievement in England. Similarly, mixed results were found in the literature, with some studies finding a difference in favour of female students (Ma, 2008), while others found none (Else-Quest, 2010). There is also a gender difference in educational aspiration in favour of female students in all three countries, similar to the findings in the literature (Schoon and Polek, 2011). In the relationship between gender and EVT factors, Japanese and English male students tend to have higher MIV, MSC, and MUV than female students, which is also in line with some literature (Eccles and Wigfield, 2002; Guo et al., 2015; Marsh et al., 2005, 2013; Steinmayr and Spinath, 2008; Watt, 2004). A more balanced result was found for Türkiye, with a trend in favour of boys in MSC and in favour of girls in MUV, with no gender difference found in MIV. Considering the mediating effects of EVT constructs between gender and achievement, these mediate in favour of male students, especially in MSC, and reduce the gender effect in Japan and Türkiye in terms of the total effect. Moreover, although the EVT constructs also mediate the relationship between gender and educational aspirations in favour of male students, the total effect for educational aspirations is still in favour of female students for all countries (see section 6.4.2 in chapter 6).

In international research on gender inequality, gender stratification is a hypothesis proposed by Else-Quest et al. (2010). It suggests that systems of wider gender inequality in society influence gender differences in mathematics achievement. As Riegle-Crumb et al. (2012) points out, girls are usually aware of their gender-specific opportunities and rely more on them to succeed. This affects their performance by affecting how much they invest in their education. According to Else-Quest et al. (2010) gender differences in mathematics and science achievement are systematically correlated with wider societal gender inequalities. Typically, scientists have historically been portrayed as men in white lab coats, shaping the image of scientists (Eccles, 1989). Since society typically considers STEM fields to be masculine, women can find it challenging to participate in them (Goldman and Penner, 2016). Therefore, these cultural stereotypes of gender may cause girls to be less motivated and feel less confident in mathematics. This is also in line with the EVT literature (Eccles, 2009; Eccles et al., 1983), which suggest that mathematics-related motivation/beliefs are more important for boys.

7.5. What is the relationship between home educational resources (HER), mathematics achievement, and educational aspirations? Do EVT factors mediate this relationship?

This research question examines the mediating effects of EVT constructs on the relationship between HER and achievement and educational aspiration through a mediation model. A moderated meditational model was developed to answer this research question. The results of this model (M16) are provided in Tables 6.1, 6.2, and 6.3 in Chapter 6 and the graph showing the structural relationship is provided in Figures 6.2, 6.3, and 6.4 for Japan, Türkiye, and England, respectively.

As hypothesised, HER have a significant positive relationship with maths achievement and educational aspirations in all three countries. In addition, HER has a significant and positive effect on all EVT structures in Japan, with MSC and MUV in Türkiye, and only MSC in England. The latter has a positive and significant mediating effect on the relationship of EVT constructs with HER and mathematics achievement and educational aspirations in all three countries. On the other hand, MIV has a negative mediating role with HER in mathematics achievement and educational aspirations in Japan. In addition, MUV has a small but significant mediating effect between HER and educational aspirations in Türkiye and Japan. In parallel with the literature and as in other models in this study, MSC is the EVT construct that has the most important effect on both the direct relationship with and the relationship between HER and educational aspirations, which is an indirect effect and educational outcomes (Guo et al., 2015; Parker et al., 2012).

This study shows that MSC mediates the relationship between HER and maths achievement in all three countries. The HER variable used as an SES indicator in this study consists of three main components: the number of books and study supports in the home and their parents' education level (see section 3.2.4 chapter 3). Family socialisation models (Eccles and Davis-Kean, 2005) have been proposed to explain the association between HER and MSC by considering certain sociocultural factors, such as socioeconomic status (SES), that impact motivational factors. Accordingly, it is claimed that parents with higher educational levels and higher SES will provide better educational experiences, set high expectations for future education paths and career plans, and act as role models for their children. Students' perceptions of their parents' expectations can play a part in determining their own expectations and task values. This may result in students modelling, communicating expectations, and providing differential experiences on the part of their parents. It is expected that students with a higher level of HER tend to perform better since this increases their self-concept. Therefore, the expectation that positive relations with HER and educational outcomes would be mediated by expectations (MSC) is justified.

The findings from RQ4 and RQ5 are strongly in line with EVT that claims expectancy (MSC) and task values (MIV and MUV) predict educational outcomes and are affected by demographic variables such as HER and gender (Eccles, 2007; Guo et al., 2015; Brown and Putwain, 2022).

7.6. Summary

This section summarises the highlights of the discussion chapter according to the research questions.

Research question 1

- The omega and alpha indices of TIMSS 2019 found motivational structures within acceptable ranges in Japan, Türkiye, and England. It should be noted, however, that negatively coded items had an adverse effect on reliability in the omega analysis. This study concludes that omega reliability indices provide a more consistent and accurate result than alpha for unidimensional or multidimensional constructs, especially for constructs that include negative items since the alpha may estimates overestimate reliability (Raykov and Shrout, 2002).
- The literature has reported negatively coded items effect, particularly in young participants (Marsh, 1996; Michaelides, 2016). This may be related to the fact that reading skills and comprehension are lower among young participants due to cognitive development. Therefore, this study suggests that the error terms of negative items in the model should be correlated to avoid such adverse effects (Marsh et al., 2013a).

- ESEM models performed better than CFA models regarding model fit values. This is because ESEM allows items to be cross loaded between constructs, which should be theoretically related (Marsh et al., 2020). ESEM enables such theoretical relationship to be applied to the statistical model with a more flexible approach than CFA.

Research question 2

- Motivational beliefs are highest in Türkiye and lowest in Japan. On the other hand, Japan has the highest mean achievement score in mathematics, while Türkiye has the lowest. These findings are "perplexing" and "paradoxical", but they are consistent with those of other large-scale studies (Bofah and Hannula, 2015; Marsh et al., 2013a; Shen and Tam, 2008). The reason may be that Japanese and Turkish students compare themselves to their classmates and classmates around them rather than students from other countries (Shen and Tam, 2008). As a result, they may have low academic expectations in low-performing countries and high academic expectations in high-performing ones.
- In line with other studies using TIMSS data, this study found a high correlation between MIV and MSC, which may compromise discriminant validity (Bofah and Hannula, 2015; Guo et al., 2015; Marsh et al., 2013a). There is, however, a significant difference in the correlation of MIV and MSC with mathematics achievement; therefore, discriminant validity is supported (Marsh et al., 2013a). Nevertheless, multicollinearity may occur in regression analyses. This should be considered when analysing these variables due to the high correlation between MSC and MIV.

Research question 3

- Utility value is a stronger predictor of mathematics achievement than intrinsic value in Japan and Türkiye. However, in England, utility value does not play a significant role in predicting mathematics achievement. Cultural differences might explain this result. In Asian cultures, collectivist cultures and corresponding school practices shape motivational perspectives, whereas performance and achievement are significantly influenced by external factors rather than personal interests and enjoyment. Leung (2001) argues that teachers in Western countries place a high value on intrinsic motivation, while educators in East Asian countries place a high value on extrinsic motivation. This study supports these arguments.
- The interaction effect between MSC and task value significantly predicts mathematics achievement. This suggests that increasing MSC and task value values simultaneously would be beneficial rather than focusing on only one motivational construct. Furthermore, the importance of MSC should be emphasised, as this and relevant literature show that a low MSC combined with a high MIV can negatively affect students' achievement (Guo et al., 2015).

Research question 4

- Even though there is no statistically significant direct relationship between gender and mathematics achievement, motivational constructs mediate this relationship in favour of males. This study found that male students have higher motivational beliefs than female students, which is consistent with the literature (Guo et al., 2015; Marsh et al., 2013a; Pampaka et al., 2011; Wigfield and Eccles, 2002). Several reasons are discussed in the literature, including the possibility that gender inequality in society may contribute to this (Else-Quest et al., 2010).

Research question 5

- It has been found that HER is significantly associated with mathematics achievement. Furthermore, mainly MSC, motivational constructs mediate the relationship between HER and mathematics achievement. There is evidence from the EVT (Eccles and Davis-Kean, 2005) that shows both the level of education in the family and the home opportunities have a positive impact on both the school achievement of the student and their self-concept and motivation to perform well.

In this chapter, I discussed the findings from the analysis results with the relevant literature and theory. A summary of key findings and their implications for practice and theory are discussed in the following conclusions chapter. The chapter also provides limitations of the study and some recommendations for future research.

Chapter 8 Conclusion

This thesis's main findings were discussed in the previous chapter. The main conclusions of this study are now presented in this chapter as a result of that discussion. As a first step in this conclusion chapter, I briefly comment on the context of the research. Then, I describe how these objectives are accomplished in this thesis, through a brief summarisation of the major findings (covered in more detail in Chapter 7). This is followed by limitation of the study. After that, implications for policy and practice, and a summary of the overall contribution of this study is presented. Next, I conclude with recommendations for future researchers interested in the relationship between students' educational outcomes and motivation via large-scale assessment data. Finally, the thesis ended with the section of reflection on the thesis.

8.1. Context of the Research

The topic of interest in this thesis is the association between students' motivation in the framework of EVT and educational outcomes by using TIMSS 2019 data of Japan, Türkiye, and England.

It firstly begins with the examination of psychometrics properties of TIMSS 2019 motivation constructs since, although using large-scale assessment data is common, surprisingly little research is interested in evaluating the validity or reliability of data. As noted in Chapter 1, my interest in this topic is to bring substantive-methodological synergy into my research, the importance of which is emphasised by Marsh et al. (2004) who highlight that a robust theoretical and substantive interpretation can be achieved using appropriate methods.

Secondly, there are discussions on the relative absence in the recent literature of expectancy-value interaction, which has an important place in the original EVT but not in the modern EVT. In this study, a small but significant effect of the interactive effect was found (see chapter 5). My stance on this issue is that it is not surprising to obtain a small effect due to the use of cross-sectional data in the study as stated by Trautwein et al. (2012). However, I believe that direction is more important than effect size in examining the interaction effect.

Third, the student's background/demographic indicators play an important role in the complex relationship between motivation and educational outcomes. Considering this motivation, Chapter 6 examined EVT as a mediating factor between demographic indicators and educational outcomes. In the discussion chapter, I have suggested some possible cultural determinants of this gender gap in motivation. However, more evidence to explain why it exists is still needed, so more research is required to identify these factors.

8.2. Major Findings and Contributions of the Study

In summary, the thesis has answered the research questions and tested the hypotheses outlined in Chapter 3 to achieve its aims. As a result, this thesis has contributed significantly to the methodological issues surrounding TIMSS motivation data and understanding of students' motivation in relation to educational outcomes (mathematics achievement and educational aspirations) and demographic factors (gender and HER). These contributions are summarised below in detail:

(1) The TIMSS 2019 motivation constructs provide overall valid and reliable measurements despite some psychometric shortcomings.

- In each of the three countries, negatively coded items in the MSC and MIV measures depress the omega reliability estimate and weaken the model fit indices (RMSEA, CFI, and TLI) considered in this research). Therefore, this detrimental effect should be controlled by allowing the correlation of the error terms of negative items (Marsh et al., 2013; Raykov, 2001). (see section 4.3.3.1 in chapter 4)
- Confirmatory factor analysis (CFA) and exploratory structural equation modelling (ESEM) were compared as measurement models. It was determined that ESEM was superior to CFA based on the model fit indices (see section 4.3.3 in chapter 4).
- Since country comparisons are made in this study, measurement invariance tests are applied. As a result, while configural and metric invariance were achieved, scalar invariance was not. Therefore, only partial scalar invariance was obtained by estimating the intercepts of some items free (see section 4.3.3.3 in chapter 4).
- There is a high correlation between MIV and MSC in all three countries, which might impact on discriminant validity. However, the correlation of these measures with mathematics achievement and educational aspiration has different effects and thus does suggest discriminant validity (see section 4.3.4.2 in chapter 4).

(2) Expectancy-value constructs predict mathematics achievement and educational aspirations in all three of the countries considered.

- When the task values (MIV and MUV) are considered alone (without MSC) (model 12), MIV is a strong predictor of mathematics achievement, while

MUV is a strong predictor of educational aspirations. However, when the expectancy variable (MSC) is included in the model, MSC becomes by far the strongest predictor for achievement and MIV loses its predictive effect (see section 5.4.1 and 5.4.2 in chapter 5).

- Expectancy-value interaction has a significant effect on maths achievement and educational aspiration. Although there are some specific differences between the three countries, in summary, a high MSC has a generally positive effect on educational outcomes. In other words, when MSC is low, an increase in MIV and MUV can have a detrimental effect on these outcomes (see section 5.4.1 and 5.4.2 in chapter 5).

(3) The relationship between demographic indicators (gender and HER) and educational outcomes (mathematics achievement and educational aspirations) is mediated by EVT structures.

- It is shown that female students are more likely to have higher mathematics achievement and educational aspirations. EVT values are higher for male students, especially in MSC. This allows EVT factors to mediate the relationship between gender and achievement and to close the gap in mathematics achievement in terms of total effects. Specifically, if males and females have equal levels of belief in EVT, female students are more likely to do well in maths. In the big picture, however, boys have a higher belief in EVT than girls, which results in equal maths achievement (see section 6.4 in chapter 6).
- Home educational resources (HER), a proxy for socio-economic status, positively predicts maths achievement and educational aspirations. It also has

a strong relationship with MSC. Therefore, this study find that HER has a direct and indirect effect on educational outcomes by promoting expectancy beliefs (see section 6.4 in chapter 6).

8.3. Limitations of this Research

Some limitations of this study are listed below.

- First, this study consists of Japanese, Turkish, and English students who participated in the TIMSS 2019 programme. The study is limited to these three countries in its capacity to analyse TIMSS motivational constructs from a psychometric perspective and explain their relationship with achievement. However, in this context, especially these three countries from different cultures rather than countries with similar cultural characteristics were analysed so that the psychometric characteristics of TIMSS constructs in different cultures and the relationship between motivation and achievement of students from different cultures could be examined.
- This study analysed a cross-sectional data set, and so fundamentally is correlational research. Structural equation modelling and regression analysis assume causal directions, however, for causal relationships, longitudinal data derived from controlled experiments are generally required (Nagengast et al., 2011). An analysis of cross-sectional data can provide insight into the associations between variables; however, the interpretation of the effect should be treated with caution when causality is asserted. This limitation might partly be compensated by using the theoretical framework developed based on earlier research evidence (Guo et al., 2015; Nagengast et al., 2011; Trautwein et al., 2012).

- In this study, we are not able take into account previous mathematics achievements as TIMSS 2019 does not provide them. There is significant evidence that previous achievement affects the learning outcomes of students (Hattie, 2012). In this case, controlling for previous mathematics achievement might have give different results.
- In TIMSS, HER is used as an SES indicator. However, this SES variable is narrowly defined and missing some arguably essential elements such as parents' income and occupation (Guo et al., 2015).
- In this study, EVT constructs, and achievement are mathematics-specific variables, but the educational aspirations variable focuses on the student's overall future educational goal with a single item. The fact that expectancies and task values differ considerably across domains makes it necessary to include items that assess a student's intentions regarding studying or taking up a career related to mathematics (Eccles & Wigfield, 2002). This is because when it comes to domains such as mathematics, a student's expectancy or task values may be an important factor in their decision to pursue a career in that field. Thus, it is important to assess these intentions in order to get a better understanding of a student's motivations for studying or taking up a career related to mathematics.
- In this study, the relationship between mathematics, educational aspiration and motivation was analysed within the scope of EVT. However, it is important to note that EVT examines the task values in four facets: attainment value, intrinsic value, utility value, and cost (Eccles and Wigfield, 2002). Therefore, this is study is limited to address all of components of task values given in EVT. Accordingly, attainment value and cost were not taken into account in

this study since there are only structures related to intrinsic and utility value in TIMSS 2019.

8.4. Recommendations and Implications

This section briefly proposes some possible recommendations and implications for teachers and policy makers, TIMSS and researchers based on the findings of this study, its limitations, and recent improvements in data availability.

8.4.1. Recommendations and Implications for teachers and policy makers

This study has practical and theoretical implications for a range of stakeholders. In this study, it was found that EVT constructs positively affected mathematics achievement and educational aspiration. In line with the literature, intrinsic value is a stronger predictor of achievement than utility value in England and Japan. However, in Türkiye, the utility value is at least as strong a predictor of achievement as the intrinsic value. The reason for this may be related to the idea that STEM fields, such as mathematics, provide more stable income (see Chapter 7 for a detailed discussion). Furthermore, MSC is the strongest predictor among the EVT constructs. One of the key contributions of this study is to investigate the effect of the interaction of expectancy and task value, which is not used in modern EVT, on mathematics achievement and educational aspiration. In this study, the interaction of MSC and MIV successfully predicts mathematics achievement. Therefore, this study provides empirical evidence that the interaction term is a critical component of EVT and provides empirical evidence to support this claim.

In conjunction with previous research, our findings can be discussed in detail with regard to practical applications for education. Our results suggest that both expectations and value beliefs should be improved simultaneously, in accordance with the findings of Guo et al. (2015) and Meyer et al. (2020). Increasing value beliefs alone may, however, have a detrimental effect on academic performance for these students while maintaining low expectations beliefs. In this manner, practitioners' attempts to increase students' expectancy-value by giving priority to MSC will positively affect academic achievement. For example, an interactive teaching and learning approach and fostering achievement motivation are among the important strategies and methodologies to enhance student motivation (De la Fuente and Justicia, 2007).

Despite the fact that there is little evidence of gender differences in mathematics achievement, the gender stereotypic difference between boys and girls in the perception of expectancy and task value in mathematics remains in favour of boys. Due to these gender differences, it is likely that girls will be underrepresented in maths-related fields, which is an important concern as a result of these differences (Parker et al., 2012). This may also cause female students to subconsciously consider themselves inadequate in STEM fields, as the typical scientist portrait is drawn over a male figure, as stated in the discussion section. Therefore, practitioners and policy developers need to implement practices that increase the self-confidence and motivation of female students in STEM fields.

The findings of this study indicate that SES has an impact on achievement, educational aspirations, and motivation. In this context, the reduction of disparities between students will have a positive effect on educational outcomes, both directly and indirectly.

8.4.2. Recommendations and Implications for TIMSS

This section provides some recommendations and implications for TIMSS. A substantial number of findings regarding construct validity have been found in the

TIMSS 2019 mathematics motivation scales. Mathematics intrinsic value, mathematics utility value, and mathematics self-concept were all moderately to strongly correlated with each other. Motivation theories consider three constructs different in line with the multidimensional conceptualisation of academic motivation (Eccles et al., 1983; Raci and Dyan, 1987). In this study, the strongest relationships were found between MSC and MIV, which is consistent with the findings reported by Marsh et al. (2013a). The high correlation between motivation constructs may seem to threaten construct validity; however, it is critical to note that the significantly different correlation between mathematics achievement and motivation constructs is evidence that construct validity is present. Nevertheless, this high correlation can potentially lead to multicollinearity in regression analyses. Therefore, the results of this study suggest that it would be helpful for the TIMSS team to revisit the items causing this high correlation and to revise these constructs. This high correlation may also be attributed to the lack of a theoretical framework in the construction of TIMSS motivational constructs. Therefore, TIMSS constructs may be improved with a more explicit theoretical framework, which may reduce their high correlation.

As another possible implication of this study, future revisions of TIMSS background questionnaires may reconsider the use of negatively stated items. The results of this study indicate that negatively coded items have a significant impact on model fit parameters. In this study, the degree of model fit is significantly reduced when the negative item effect is not controlled, which was accomplished by correlating the error terms. The use of positive statements is an alternative to using negative statements; however, the intended benefit of reducing response sets would be lost if the statements were reworded to use positive statements (Michaelides, 2019). As another option, they could be kept in the instrument, but instead of counting them for scoring purposes, they could just be retained for the instrument’s purposes (Marsh, 1996). In other words, negatively worded items can be kept in the instruments because they can lead to more thoughtful responses from the participants, which can help to reduce the chance of them responding in an acquiescent manner. In order to prevent problems with model fit, negatively worded items can be removed before scoring. At the very least, TIMSS should consider this situation's detrimental impact when converting constructs with negative items into manifest scores.

In the TIMSS 2019 assessment, the variable HER is used as a student socioeconomic indicator. This data includes the educational levels of parents and the number of books in the home. This can be considered a narrow definition of the SES indicator (Guo et al., 2015). Therefore, if the TIMSS committee were to provide additional information about the family’s occupation and income for this variable used as an SES indicator, it would make it a more comprehensive indicator. Furthermore, as mentioned in Chapter 2, even though self-determination theory and self-concept theories are referred to in the construction of TIMSS motivational constructs, the theoretical frameworks of these constructs are not clearly stated. I believe that a review of these constructs in TIMSS in a clearer theoretical framework will contribute positively to the validity and reliability of the motivational constructs in terms of theory and for researchers to adapt these constructs to their own studies.

There is a need to develop items that will assess students’ intention to study or pursue a career related to maths, as both expectancy and task values are highly domain specific (Eccles and Wigfield, 2002). It is important to note that the EVT examines the task values in four facets: attainment value, intrinsic value, utility value, and cost (Eccles and Wigfield, 2002). However, attainment value and cost were not considered in this study since there are only structures related to intrinsic and utility value in TIMSS 2019.

8.4.3. Recommendations for researchers

This study uses structural equation modelling (SEM) as a statistical analysis technique. It would be helpful for the investigation to experiment with different alternative statistical approaches that address relatively recent issues (e.g., Rash modelling or multilevel modelling). In this context, it should be reminded that the SEM method has many advantages over the traditional regression method, such as reducing measurement error and negative item effect. As mentioned in the previous section, negative items have a significant effect on model fit and model parameters. Therefore, this study encourages researchers to use latent variable models to reduce the effect of these negative items.

As stated in Section 8.4, this study used cross-sectional data. Future studies could replicate this study with longitudinal data with a suitable data set. In this way, a more robust conclusion about interaction effects would be obtained. In addition, in this study, only two subcomponents of task value, intrinsic and utility, have been invented. Future studies may investigate other components of task value (attainment value and cost) where the data are available.

In this study, gender and socioeconomic factors as background/demographic variables are applied. In future studies, research with different variables that may have a direct effect on the motivation and achievement of the student, such as the expectations of the family and the teacher, will contribute to understanding this complex relationship among students' motivation and educational outcomes and demographic factors.

This study is limited to Türkiye, Japan, and England. Therefore, this study could be repeated with different countries participating in TIMSS in order to evaluate TIMSS 2019 data in terms of validity and reliability and to improve the effect of motivational structures in different educational systems and cultures in both theoretical and practical terms.

8.6. Reflections on the thesis

After a long and tiring but instructive PhD process, I am the final steps away from finishing my thesis. The purpose of this section is to share my personal experiences regarding my PhD journey and the process of writing my thesis.

The scope of my thesis has significantly changed from what I envisioned at the beginning and what it has become as it has been finalised. Along the way, I encountered various challenges, such as finding the right sources, understanding the technical aspects of my thesis, and dealing with the complexity of research. I had to adjust my scope multiple times to ensure that I was addressing the most relevant topics. My doctoral studies were initially intended to examine the school, class and motivational factors influencing students' mathematics achievement from a broader perspective; however, this idea evolved into a more specific investigation of the relationship between motivation and achievement, guided by my supervisors and the literature. There were several factors that played a significant role in this transformation, including the articles I read in the literature and their methodology. In my initial observations, I observed that studies addressing the factors affecting success from a broader perspective generally used single or multilevel regression models based on manifest scores. Comparatively, I have observed that specialised methods, such as the SEM method, are more effective in studies that focus on a narrower topic, such as the present study. This is because the SEM method can provide more accurate results due to its ability to account for the complex relationships between variables. Furthermore, the SEM method allows for the inclusion of latent variables, which can help to provide a more comprehensive analysis of the data. The articles and studies that I read in the literature influenced my decision during this critical stage of the process. Specifically, Herbert Marsh's studies on motivation, self-concept, and EVT were influential in shaping my thesis which interested in both theoretical and substantive aspects of these constructs and the psychometric properties of the constructs used to measure them.

The theoretical background of the thesis developed in parallel with the clarification of its purpose and scope. Expectancy-value theory (EVT) and self-determination theory (SDT) theories are prominent in studies of student motivation. EVT and SDT provide insight into the factors that influence students' motivation and engagement, and they help to explain why students may have different levels of motivation. These theories have been used to explain how students' perceptions of their environment and their goals can impact their motivation levels. Even though both theories explain the relationship between motivation and achievement in many similar ways, some differences also exist between the two theories. The study's theoretical background was determined by these differences. In my research, EVT theories have a more appropriate structure regarding the variables in the research and the study's objectives. Furthermore, EVT more accurately reflects how different variables interact with each other, allowing for a more comprehensive understanding of the effects of motivation on students' academic performance. Ultimately, we determined that the study should be based on the EVT theory in light of the literature, as the EVT theory incorporates not only the motivational structures (intrinsic value and utility value) but also the self-concept (expectancy of success) structure in relation to achievement, and the direction in which this connection is oriented. It should be noted that the adapted theory was not selected because it was superior to other theories but rather because it was more appropriate for the purposes and scope of the research.

One of the most challenging aspects of the thesis writing process was developing a methodology and analysing the data. This thesis used a secondary dataset. Despite its many advantages, it also has some disadvantages. Perhaps the greatest advantage of secondary data is that the process of collecting the data is carried out on your behalf by a professional organization that uses proficient methodologies. There are, however, a number of disadvantages to this, such as not being able to specify a framework for surveys or questionnaires. There might be some limitations to the theoretical construction and conceptualization of questionnaire structures because the content was determined by someone else. On the other hand, the study consists of data from three different countries from a large number of students. This means it is unlikely that a doctoral student will be reached individually.

The complex data structure used in TIMSS as well as the complexity of the SEM analysis method, has been challenging for someone like me who does not have a background in advanced statistical methods. This process required me to participate in training in statistics and quantitative research methods from numerous institutions, including the University of Manchester, the University of Nottingham, and the International Association for the Evaluation of Educational Achievement (IEA). I was especially encouraged to pursue these courses by my supervisor, Matt. After the training, I started to analyse the classical measuring model used in the SEM method with CFA. After performing a more in-depth analysis, I realised that the ESEM model might be superior in such cases in terms of flexibility, accuracy, and overall capability compared to the classical CFA model. Additionally, the ESEM model was able to accommodate more complex research questions, allowing me to explore my research topic on a deeper level. Following the primary analysis method (SEM), I added a new research question to my thesis, aiming to contribute to the comparison of CFA and ESEM in a more advanced measurement model in the current literature. By introducing a more complex and realistic measurement model, I sought to provide a more nuanced comparison of CFA and ESEM.

The present thesis is the outcome of this process. My academic identity and interests were discovered during this journey, which was rewarding and instructive.

List of References

Abu-Hilal, M.M. and Aalhussain, A.-Q.A. 1997. Dimensionality and hierarchy of the sdq in a non-western milieu: A test of self-concept invariance across gender. *Journal of cross-cultural psychology.* **28**(5), pp.535-553.

Aiken, L.S. and West, S.G. 1991. *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Akben-Selcuk, E. 2017. Personality, motivation, and math achievement among turkish students:Evidence from pisa data. *Perceptual and Motor Skills.* **124**(2), pp.514-530.

Alamer, A. 2021. Construct validation of self-determination theory in second language scale: The bifactor exploratory structural equation modeling approach. *Frontiers in Psychology.* **12**.

Alamer, A. and Marsh, H. 2022. Exploratory structural equation modeling in second language research: An applied example using the dualistic model of passion. *Studies in Second Language Acquisition.* **44**(5), pp.1477-1500.

Areepattamannil, S., Freeman, J.G. and Klinger, D.A. 2011. Intrinsic motivation, extrinsic motivation, and academic achievement among indian adolescents in canada and india. *Social Psychology of Education.* **14**(3), pp.427-439.

Argün, Z., Arıkan, A., Bulut, S. and Sriraman, B. 2010. A brief history of mathematics education in turkey: K-12 mathematics curricula. *ZDM.* **42**, pp.429-441.

Asparouhov, T. and Muthén, B. 2009. *Structural Equation Modeling.* **16**, p397.

Asparouhov, T., Muthén, B. and Morin, A.J.S. 2015. Bayesian structural equation modeling with cross-loadings and residual covariances:Comments on stromeyer et al. *Journal of Management.* **41**(6), pp.1561-1577.

Atkinson, J.W. 1957. Motivational determinants of risk-taking behavior. *Psychological review.* **64**(6p1), p359.

Atkinson, J.W. and Feather, N.T. 1966. *A theory of achievement motivation.* New York: Wiley.

Bandalos, D.L. 2014. Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal.* **21**(1), pp.102-116.

Bandura, A. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological review.* **84**(2), p191.

Bandura, A. 1997. *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Battle, E.S. 1966. Motivational determinants of academic competence. *Journal of Personality and Social Psychology*. **4**(6), p634.

Becker, M., McElvany, N. and Kortenbruck, M. 2010. Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology*. **102**, pp.773-785.

Benson, J. and Hocevar, D. 1985. The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement.* **22**(3), pp.231-240.

Bentler, P.M. 2009. Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika.* **74**(1), pp.137-143.

Bofah, E.A. and Hannula, M.S. 2014. *Structural equation modelling: Testing for the factorial validity, replication and measurement invariance of students' views on mathematics views on mathematics.* SAGE.

Bofah, E.A.-t. and Hannula, M.S. 2015. Timss data in an african comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties. *Large-scale Assessments in Education.* **3**(1).

Bollen, K.A. 1989. Measurement models: The relation between latent and observed variables. *Structural equations with latent variables.* pp.179-225.

Bollen, K.A. 1989. *Structural equations with latent variables.*

Brophy, S., Klein, S., Portsmore, M. and Rogers, C. 2008. Advancing engineering education in p-12 classrooms. *Journal of Engineering Education.* **97**(3), pp.369-387.

Brown, C. and Putwain, D.W. 2022. Socio-economic status, gender and achievement: The mediating role of expectancy and subjective task value. *Educational Psychology.* **42**(6), pp.730-748.

Brown, T.A. 2015. *Confirmatory factor analysis for applied research.* 2nd ed. New York, NY: Guilford Press.

Busemeyer, J.R. and Jones, L.E. 1983. Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin.* **93**, pp.549-562.

Byrne, B.M. and Shavelson, R.J. 1987. Adolescent self-concept: Testing the assumption of equivalent structure across gender. *American Educational Research Journal.* **24**(3), pp.365-385.

Byrne, B.M., Shavelson, R.J. and Muthén, B. 1989. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin.* **105**, pp.456-466.

Carmines, E.G. and Zeller, R.A. 1979. *Reliability and validity assessment.* Sage publications.

Caro, D.H., Sandoval-Hernández, A. and Lüdtke, O. 2014. Cultural, social, and economic capital constructs in international assessments: An evaluation using exploratory structural equation modeling. *School Effectiveness and School Improvement.* **25**(3), pp.433-450.

Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. 2006. *Measurement error in nonlinear models: A modern perspective.* Chapman and Hall/CRC.

Chen, F.F. 2008. What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *J Pers Soc Psychol.* **95**(5), pp.1005-1018.

Chen, J.J.-L. 2005. Relation of academic support from parents, teachers, and peers to hong kong adolescents' academic achievement: The mediating role of academic engagement. *Genetic, social, and general psychology monographs.* **131**(2), pp.77-127.

Cheung, G.W. and Rensvold, R.B. 2001. The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods.* **4**, pp.236-264.

Chiu, M.-S. 2012. Differential psychological processes underlying the skill-development model and self-enhancement model across mathematics and science in 28 countries. *International Journal of Science and Mathematics Education.* **10**(3), pp.611-642.

Choi, K., Choi, T. and McAninch, M. 2012. A comparative investigation of the presence of psychological conditions in high achieving eighth graders from timss 2007 mathematics. *ZDM.* **44**(2), pp.189-199.

Clarke, M. and Luna-Bazaldua, D. 2021. *Primer on large-scale assessments of educational achievement.* World Bank Publications.

Conley, A.M. 2012. Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology.* **104**, pp.32-47.

d'Ailly, H. 2003. Children's autonomy and perceived control in learning: A model of motivation and achievement in taiwan. *Journal of Educational Psychology.* **95**(1), pp.84-96.

Dawes, L. and Rasmussen, G. 2007. Activity and engagement–keys in connecting engineering with secondary school students. *Australasian Journal of Engineering Education.* **13**(1), pp.13-20.

de la Fuente, J., Sander, P. and Putwain, D. 2013. Relationship between undergraduate student confidence, approach to learning and academic performance: The role of gender. *Revista de Psicodidáctica.* **18**(2), pp.375-393.

Deci, E.L. and Moller, A.C. 2005. The concept of competence: A starting place for understanding intrinsic motivation and self-determined extrinsic motivation. *Handbook of competence and motivation.* New York, NY, US: Guilford Publications, pp.579-597.

Deci, E.L. and Ryan, R.M. 1985. *Intrinsic motivation and self-determination in human behavior.* London;New York;: Plenum.

Deci, E.L. and Ryan, R.M. 2012. Self-determination theory.

Denissen, J.J.A., Zarrett, N.R. and Eccles, J.S. 2007. I like to do it, i'm able, and i know i am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development.* **78**(2), pp.430-447.

Distefano, C. and Motl, R. 2009. Methodological artifact or substance? Examinations of wording effects associated with negatively worded items. *Structural equation modeling in educational research: Concepts and applications.* pp.59-77.

Dodge, Y., Cox, D. and Commenges, D. 2003. *The oxford dictionary of statistical terms.* Oxford University Press on Demand.

Eccles, J. 2011. Gendered educational and occupational choices: Applying the eccles et al. Model of achievement-related choices. *International Journal of Behavioral Development.* **35**(3), pp.195-201.

Eccles, J.S. 1994. Understanding women's educational and occupational choices: Applying the eccles et al. Model of achievement-related choices. *Psychology of Women Quarterly.* **18**(4), pp.585-609.

Eccles, J.S. 2009. Who am i and what am i going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist.* **44**(2), pp.78-89.

Eccles, J.S., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, L.J. and al., e. 1983. Expectancies, values, and academic behaviors. In: Spence, J.T. ed. *Achievement and achievement motivation: Psychological and sociological approaches.* San Francisco, CA: Freeman, pp.75-146.

Eccles, J.S. and Wigfield, A. 1995. In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and social psychology bulletin.* **21**(3), pp.215-225.

Eccles, J.S. and Wigfield, A. 2002. Motivational beliefs, values, and goals. *Annual review of psychology.* **53**(1), pp.109-132.

Eccles, J.S. and Wigfield, A. 2020. From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary educational psychology.* **61**, p101859.

Edwards, J.R. and Bagozzi, R.P. 2000. On the nature and direction of relationships between constructs and measures. *Psychological methods.* **5**(2), p155.

Else-Quest, N.M., Hyde, J.S. and Linn, M.C. 2010. Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological bulletin.* **136**(1), p103.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. and Strahan, E.J. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods.* **4**(3), p272.

Fadda, D., Scalas, L.F., Morin, A.J., Marsh, H.W. and Gaspard, H. 2020. Value beliefs about math: A bifactor-esem representation. *European Journal of Psychological Assessment.* **36**(2), p259.

Feather, N.T. 1982. Expectancy–value approaches: Present status and future directions. In: Feather, N.T. ed. *Expectations*

and actions: Expectancy–value models in psychology. Hillsdale, NJ: Lawrence Erlbaum, pp.395-420.

Feather, N.T. 1988. Values, valences, and course enrollment: Testing the role of personal values within an expectancy-valence framework. *Journal of educational psychology.* **80**(3), p381.

Fishbein, B., Foy, P. and Yin, L. 2021. *Timss 2019 user guide for the international database.* [Online]. [Accessed 25 February 2023]. Available from: https://timssandpirls.bc.edu/timss2019/international-database/

Frenzel, A.C., Pekrun, R. and Goetz, T. 2007. Girls and mathematics —a "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education.* **22**(4), p497.

Fu, Y., Wen, Z. and Wang, Y. 2022. A comparison of reliability estimation based on confirmatory factor analysis and exploratory structural equation models. *Educational and Psychological Measurement.* **82**(2), pp.205-224.

Furnham, A., Guenole, N., Levine, S.Z. and Chamorro-Premuzic, T. 2013. The neo personality inventory–revised:Factor structure and gender invariance from exploratory structural equation modeling analyses in a high-stakes setting. *Assessment.* **20**(1), pp.14-23.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U. and Nagengast, B. 2015. More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology.* **107**, pp.663-677.

Goetz, T., Frenzel, A.C., Hall, N.C. and Pekrun, R. 2008. Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology.* **33**(1), pp.9-33.

Goldman, A.D. and Penner, A.M. 2016. Exploring international gender differences in mathematics self-concept. *Int J Adolesc Youth.* **21**(4), pp.403-418.

Gomez, R., Stavropoulos, V. and Griffiths, M.D. 2020. Confirmatory factor analysis and exploratory structural equation modelling of the factor structure of the depression anxiety and stress scales-21. *PLOS ONE.* **15**(6), pe0233998.

Greaney, V. and Kellaghan, T. 2008. *Assessing national achievement levels in education.* World Bank Publications.

Grolnick, W.S., Friendly, R.W. and Bellas, V.M. 2009. Parenting and children's motivation at school. *Handbook of motivation at school.* New York, NY, US: Routledge/Taylor & Francis Group, pp.279-300.

Grønmo, L.S., Lindquist, M., Arora, A. and Mullis, I.V. 2015. *Timss 2015 mathematics framework.* [Online]. [Accessed 25 February 2023]. Available from: https://timssandpirls.bc.edu/timss2015/

Guay, F., Larose, S. and Boivin, M. 2004. Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity.* **3**, pp.53-68.

Guay, F., Morin, A.J.S., Litalien, D., Valois, P. and Vallerand, R.J. 2015. Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of experimental education.* **83**(1), pp.51-82.

Guay, F., Ratelle, C.F., Roy, A. and Litalien, D. 2010. Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences.* **20**, pp.644-653.

Guo, J. 2016. *The relations among academic motivation, self-concept, aspirations and choices: Integrating expectancy-value and academic self-concept theory*. thesis, Australian Catholic University.

Guo, J., Marsh, H.W., Parker, P.D., Dicke, T. and Van Zanden, B. 2019. Countries, parental occupation, and girls' interest in science. *Lancet.* **393**(10171), pp.e6-e8.

Guo, J., Marsh, H.W., Parker, P.D., Morin, A.J.S. and Yeung, A.S. 2015. Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences.* **37**, pp.161-168.

Guo, J., Nagengast, B., Marsh, H.W., Kelava, A., Gaspard, H., Brandt, H., Cambria, J., Flunger, B., Dicke, A.-L., Häfner, I., Brisson, B. and Trautwein, U. 2016. Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open.* **2**(1).

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. 2019. *Multivariate data analysis.* Eighth edition ed. Australia: Cengage.

Halle, T.G., Kurtz-Costes, B. and Mahoney, J.L. 1997. Family influences on school achievement in low-income, african american children. *Journal of Educational Psychology.* **89**, pp.527-537.

Haneuse, S. and Bartell, S. 2011. Designs for the combination of group- and individual-level data. *Epidemiology.* **22**(3), pp.382-389.

Harter, S. 1978. Effectance motivation reconsidered: Toward a developmental model. *Human Development.* **21**, pp.34-64.

Harter, S. 1999. *The construction of the self: A developmental perspective.* New York, NY, US: Guilford Press.

Hattie, J. 1992. *Self-concept.* Psychology Press.

Hattie, J. 2008. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* routledge.

Hau, K.T. and Marsh, H.W. 2004. The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *Br J Math Stat Psychol.* **57**(Pt 2), pp.327-351.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M. and Bühner, M. 2011. Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychol Methods.* **16**(3), pp.319-336.

Heine, S.J. 2001. Self as cultural product: An examination of east asian and north american selves. *Journal of personality.* **69**(6), pp.881-905.

Hernandez, M. and Iyengar, S.S. 2001. What drives whom? A cultural perspective on human agency. *Social Cognition.* **19**(3), pp.269-294.

Hooper, M., S., M.I.V. and Martin, M.O. 2013. Timss 2015 context questionnaire framework. In: Mullis, I.V.S. and Martin, M.O. eds. *Timss 2015 assessment frameworks.* Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College, pp.61-82.

Hooper, M., S., M.I.V., Martin, M.O. and Fishbein, B. 2017. Timss 2019 context questionnaire framework. In: Mullis, I.V.S. and Martin, M.O. eds. *Timss 2019 assessment frameworks.* Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College, pp.57-58.

Hou, K.K. 2006. *Fundamental law of education.*

House, J.D. 2008. Science beliefs, instructional strategies, and life sciences achievement of adolescent students in japan: Results from the timss 1999 assessment. *International Journal of Instructional Media.* **35**(1), pp.103-114.

Hoyle, R.H. ed. 2012. *Handbook of structural equation modeling.* New York, NY, US: The Guilford Press.

Hu, L.t. and Bentler, P.M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal.* **6**(1), pp.1-55.

Hyde, J.S., Fennema, E. and Lamon, S.J. 1990. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin.* **107**, pp.139-155.

Isaacs, T., Golding, J., Richardson, M. and Swensson, C. 2020. England. In: Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. ed. *Timss 2019 encyclopedia: Education policy and curriculum in mathematics and science.* Boston College.

Jacobs, J.E., Lanza, S., Osgood, D.W., Eccles, J.S. and Wigfield, A. 2002. Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development.* **73**(2), pp.509-527.

Jöreskog, K.G., Golberger, A.S. and Duncan, O.D. 1973. *Structural equation models in the social sciences.*

Jöreskog, K.G., Yang, F., Marcoulides, G. and Schumacker, R. 1996. Nonlinear structural equation models: The kenny-judd model with interaction effects. *Advanced structural equation modeling: Issues and techniques.* **3**, pp.57-88.

Joshanloo, M. and Lamers, S.M. 2016. Reinvestigation of the factor structure of the mhc-sf in the netherlands: Contributions of exploratory structural equation modeling. *Personality and individual differences.* **97**, pp.8-12.

Jung, J.Y. 2019. *A comparison of cfa and esem approaches using timss science attitudes items: Evidence from factor structure and measurement invariance*. thesis, Purdue University Graduate School.

Kahn, J.H. 2006. Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The counseling psychologist.* **34**(5), pp.684-718.

Kaplan, O. 2018. *Analyzing the effects of motivation, expectations and self-concept on eighth-grade students' mathematics achievement*. PhD thesis, University of Florida

Kenny, D.A. and Judd, C.M. 1984. Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin.* **96**, pp.201-210.

Ker, H. 2013. Trend analysis on mathematics achievements: A comparative study using timss data. *Universal Journal of Educational Research.* **1**(3), pp.200-203.

Ker, H.-W. 2017. The effects of motivational constructs and engagements on mathematics achievements: A comparative study using timss 2011 data of chinese taipei, singapore, and the USA. *Asia Pacific Journal of Education.* **37**(2), pp.135-149.

Kessels, U. and Hannover, B. 2008. When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology.* **78**(2), pp.273-289.

Kessels, U., Heyder, A., Latsch, M. and Hannover, B. 2014. How gender differences in academic engagement relate to students' gender identity. *Educational Research.* **56**(2), pp.220-229.

Klein, A. and Moosbrugger, H. 2000. Maximum likelihood estimation of latent interaction effects with the lms method. *Psychometrika.* **65**(4), pp.457-474.

Kline, R.B. 2015. *Principles and practice of structural equation modeling.* New York: Guilford Publications.

Kocur, D., Flakus, M. and Fopka-Kowalczyk, M. 2022. Validity and reliability of the polish version of the self-compassion scale and its correlates. *Plos one.* **17**(5), pe0267637.

Kriegbaum, K. and Spinath, B. 2016. Explaining social disparities in mathematical achievement: The role of motivation. *European Journal of Personality.* **30**(1), pp.45-63.

Kupari, P. and Nissinen, K. 2013. Background factors behind mathematics achievement in finnish education context: Explanatory models based on timss 1999 and timss 2011 data. In: *IEA CONFERENCE 2013, Proceedings*: Citeseer.

LaRoche, S., Joncas, M. and Foy, P. 2020. Sample design in timss 2019. In: M. O. Martin, et al. eds. *Methods and procedures: Timss 2019 technical report.* Boston College: TIMSS & PIRLS International Study Center.

Lauermann, F., Tsai, Y.M. and Eccles, J.S. 2017. Math-related career aspirations and choices within eccles et al.'S expectancy-value theory of achievement-related behaviors. *Dev Psychol.* **53**(8), pp.1540-1559.

Leung, F.K.S. 2001. In search of an east asian identity in mathematics education. *Educational Studies in Mathematics.* **47**(1), pp.35-51.

Li, C.H. 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods.* **48**(3), pp.936-949.

Liu, Y. and Hou, S. 2018. Potential reciprocal relationship between motivation and achievement: A longitudinal study. *School Psychology International.* **39**(1), pp.38-55.

Luttrell, V.R., Callen, B.W., Allen, C.S., Wood, M.D., Deeds, D.G. and Richard, D.C.S. 2010. The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement.* **70**, pp.142-160.

MacCallum, R.C. and Mar, C.M. 1995. Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin.* **118**, pp.405-421.

MacKinnon, Fairchild, A.J. and Fritz, M.S. 2007. Mediation analysis. *Annu Rev Psychol.* **58**, pp.593-614.

MacKinnon, D.P. 2008. *Introduction to statistical mediation analysis.*

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

Marks, G.N. 2013. *Education, social background and cognitive ability: The decline of the social.* Routledge.

Markus, H.R. and Kitayama, S. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review.* **98**(2), p224.

Marsh, H.W. 1986. Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology.* **22**, pp.37-49.

Marsh, H.W. 2007. *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology.*

Marsh, H.W., Abduljabbar, A.S., Abu-Hilal, M.M., Morin, A.J.S., Abdelfattah, F., Leung, K.C., Xu, M.K., Nagengast, B. and Parker, P. 2013a. Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of arab and anglo-saxon countries. *Journal of Educational Psychology.* **105**(1), pp.108-128.

Marsh, H.W., Balla, J.R. and McDonald, R.P. 1988. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin.* **103**, pp.391-410.

Marsh, H.W. and Craven, R.G. 2006. Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science.* **1**(2), pp.133-163.

Marsh, H.W., Guo, J., Dicke, T., Parker, P.D. and Craven, R.G. 2020. Confirmatory factor analysis (cfa), exploratory structural equation modeling (esem), and set-esem: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research.* **55**(1), pp.102-119.

Marsh, H.W. and Hau, K.-T. 2007. Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary educational psychology.* **32**(1), pp.151-170.

Marsh, H.W., Hau, K.-T., Artelt, C., Baumert, J. and Peschar, J.L. 2006. Oecd's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing.* **6**(4), pp.311-360.

Marsh, H.W., Hau, K.-T., Wen, Z., Nagengast, B. and Morin, A.J.S. 2013b. Moderation. *The oxford handbook of quantitative methods: Statistical analysis, vol. 2.* New York, NY, US: Oxford University Press, pp.361-386.

Marsh, H.W. and Hocevar, D. 1991. Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and teacher education.* **7**(4), pp.303-314.

Marsh, H.W. and Köller, O. 2004. Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and west german school systems after the fall of the berlin wall. *Contemporary Educational Psychology.* **29**(3), pp.264-282.

Marsh, H.W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A.J., Trautwein, U. and Nagengast, B. 2010. A new look at the big five factor structure through exploratory structural equation modeling. *Psychological assessment.* **22**(3), p471.

Marsh, H.W., Morin, A.J., Parker, P.D. and Kaur, G. 2014. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology.* **10**, pp.85-110.

Marsh, H.W., Morin, A.J.S., Parker, P.D. and Kaur, G. 2014. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology.* **10**(1), pp.85-110.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J. and Trautwein, U. 2009. Exploratory structural equation modeling, integrating cfa and efa: Application to students' evaluations of university teaching. *Structural equation modeling: A multidisciplinary journal.* **16**(3), pp.439-476.

Marsh, H.W., Nagengast, B. and Morin, A.J.S. 2013c. Measurement invariance of big-five factors over the life span: Esem tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental psychology.* **49 6**, pp.1194-1218.

Marsh, H.W., Parker, P.D., Guo, J., Basarkod, G., Niepel, C. and Van Zanden, B. 2021. Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology.* **121**, pp.168-183.

Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O. and Baumert, J. 2005. Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Dev.* **76**(2), pp.397-416.

Marsh, H.W., Van Zanden, B., Parker, P.D., Guo, J., Conigrave, J. and Seaton, M. 2019. Young women face disadvantage to enrollment in university stem coursework regardless of prior achievement and attitudes. *American Educational Research Journal.* **56**(5), pp.1629-1680.

Marsh, H.W., Wen, Z. and Hau, K.-T. 2004. Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological methods.* **9**(3), p275.

Martin, A.J. 2001. The student motivation scale: A tool for measuring and enhancing motivation. *Journal of Psychologists and Counsellors in Schools.* **11**, pp.1-20.

Martin, A.J., Marsh, H.W. and Debus, R.L. 2001a. A quadripolar need achievement representation of self-handicapping and defensive pessimism. *American Educational Research Journal.* **38**(3), pp.583-610.

Martin, A.J., Marsh, H.W. and Debus, R.L. 2001b. Self-handicapping and defensive pessimism: Exploring a model of predictors and outcomes from a self-protection perspective. *Journal of Educational Psychology.* **93**(1), p87.

Meece, Wigfield, A. and Eccles, J.S. 1990. Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of educational psychology.* **82**(1), p60.

Meece, J.L., Wigfield, A. and Eccles, J.S. 1990. Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology.* **82**, pp.60-70.

Mercier, H. and Sperber, D. 2017. *The enigma of reason.* Harvard University Press.

Metsämuuronen, J. 2012. Challenges of the fennema-sherman test in the international comparisons. *International Journal of Psychological Studies.* **4**(3), p1.

MEXT, M.o.E., Culture, Sports, Science, and Technology. 2019. *Gakkou kihon chousa.* Tokyo.

Meyer, J., Fleckenstein, J. and Köller, O. 2019. Expectancy value interactions and academic achievement: Differential relationships with achievement measures. *Contemporary Educational Psychology.* **58**, pp.58-74.

Michaelides, M.P. 2019. Negative keying effects in the factor structure of timss 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education.* **32**(4), pp.365-378.

Michaelides, M.P., Brown, G.T.L., Eklöf, H. and Papanastasiou, E.C. 2019. The relationship of motivation with achievement in mathematics. In: Michaelides, M.P., et al. eds. *Motivational profiles in timss mathematics: Exploring student clusters across countries and time.* Cham: Springer International Publishing, pp.9-23.

Millsap, R.E. and Olivera-Aguilar, M. 2012. Investigating measurement invariance using confirmatory factor analysis. *Handbook of structural equation modeling.* New York, NY, US: The Guilford Press, pp.380-392.

Minkov, M. 2008. Self-enhancement and self-stability predict school achievement at the national level. *Cross-Cultural Research.* **42**(2), pp.172-196.

Mislevy, R.J. 1991. Randomization-based inference about latent variables from complex samples. *Psychometrika.* **56**(2), pp.177-196.

Mone milli eğitim bakanlığının teşkilat ve görevleri hakkında kanun hükmünde kararname [decree law regarding the organizational chart and responsibilities of the ministry of national education] 2011.

MoNE. 2017. *Mone türk eğitim sistemi [turkish education system].*

Monseur, C. and Adams, R. 2009. Plausible values: How to deal with their limitations. *J Appl Meas.* **10**(3), pp.320-334.

Morin, A.J., Myers, N.D. and Lee, S. 2020. Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (esem), and bifactor-esem. *Handbook of sport psychology.* pp.1044-1073.

Morin, A.J.S., Arens, A.K. and Marsh, H.W. 2016. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal.* **23**(1), pp.116-139.

Morin, A.J.S., Marsh, H.W. and Nagengast, B. 2013. Exploratory structural equation modeling. *Structural equation modeling: A second course, 2nd ed.* Charlotte, NC, US: IAP Information Age Publishing, pp.395-436.

Mullis, I.V., Martin, M.O., Foy, P. and Arora, A. 2012. *Timss 2011 international results in mathematics.* ERIC.

Mullis, I.V., Martin, M.O., Foy, P., Kelly, D.L. and Fishbein, B. 2020. *Timss 2019 international results in mathematics and science.* [Online]. Available from: https://timssandpirls.bc.edu/timss2019/international-results/

Mullis, I.V.S., Martin, M.O., Foy, P. and Hooper, M. 2016. *International results in mathematics.* Boston College: TIMSS & PIRLS International Study Center.

Murayama, K., Pekrun, R., Lichtenfeld, S. and vom Hofe, R. 2013. Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development.* **84**, pp.1475-1490.

Musu-Gillette, L.E., Wigfield, A., Harring, J.R. and Eccles, J.S. 2015. Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational research and evaluation.* **21**(4), pp.343-370.

Muthen, B.O. 2015. Test for moderating variables. [Online]. Available from: http://www.statmodel.com/discussion/messages/11/69.html

Muthén, L.K. and Muthén, B.O. 1998-2017. *Mplus user's guide.* Eighth ed. Los Angeles: CA: Muthén & Muthén.

Nagengast, B., Marsh, H.W., Scalas, L.F., Xu, M.K., Hau, K.T. and Trautwein, U. 2011. Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychol Sci.* **22**(8), pp.1058-1066.

Nagengast, B., Trautwein, U., Kelava, A. and Lüdtke, O. 2013. Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research.* **48**, pp.428-460.

Nagy, G., Garrett, J., Trautwein, U., Cortina, K.S., Baumert, J. and Eccles, J.S. 2008. Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences.* Washington, DC, US: American Psychological Association, pp.115-143.

Nagy, G., Trautwein, U., Baumert, J., Köller, O. and Garrett, J. 2006. Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational research and Evaluation.* **12**(4), pp.323-345.

Novick, M.R. and Lewis, C. 1967. Coefficient alpha and the reliability of composite measurements. *Psychometrika.* **32**(1), pp.1-13.

OECD. 2003. *Pisa literary skills for the world of tomorrow—further results from pisa.* Paris, France: Organisation for Economic Cooperation and Development.

OECD. 2013. *Pisa 2012 results: What students know and can do. Student performance in mathematics, reading and science.* OECD Publishing: Organisation for Economic Cooperation and Development.

OECD. 2019. *Pisa 2018 assessment and analytical framework.*

Pampaka, M., Kleanthous, I., Hutcheson, G.D. and Wake, G. 2011. Measuring mathematics self-efficacy as a learning outcome. *Research in Mathematics Education.* **13**(2), pp.169-190.

Parker, P.D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U. and Eccles, J.S. 2012. Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology.* **48**, pp.1629-1642.

Parlak, B., Yildirim, A. and Özdemir, E. 2019. *Turkey.* [Online]. [Accessed 25 February 2023]. Available from: https://timssandpirls.bc.edu/timss2019/encyclopedia/turkey.html

Ping Jr, R.A. 1995. A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research.* **32**, pp.336-347.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y. and Podsakoff, N.P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology.* **88**, pp.879-903.

Raykov, T. 2001. Bias of coefficient afor fixed congeneric measures with correlated errors. *Applied Psychological Measurement.* **25**(1), pp.69-76.

Raykov, T. 2012. Scale construction and development using structural equation modeling. *Handbook of structural equation modeling.* New York, NY, US: The Guilford Press, pp.472-492.

Raykov, T. and Shrout, P.E. 2002. Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal.* **9**(2), pp.195-212.

Richardson, M., Isaacs, T., Barnes, I., Swensson, C., Wilkinson, D. and Golding, J. 2020. Trends in international mathematics and science study (timss) 2019: National report for england: Research report: December 2020.

Riegle-Crumb, C., King, B., Grodsky, E. and Muller, C. 2012. The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into stem college majors over time. *American Educational Research Journal.* **49**(6), pp.1048-1073.

Roldán, J.L. and Sánchez-Franco, M.J. 2012. Variance-based structural equation modeling: Guidelines for using partial least squares in information systems research. In: Mora, M., et al. eds. *Research methodologies, innovations and philosophies in software systems engineering and information systems.* Hershey, PA, USA: IGI Global, pp.193-221.

Rosenberg, M. 1979. *Conceiving the self.* United States.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons.

Rutkowski, L., Gonzalez, E., Joncas, M. and von Davier, M. 2010. International large-scale assessment data:Issues in secondary analysis and reporting. *Educational Researcher.* **39**(2), pp.142-151.

Rutkowski, L. and Rutkowski, D. 2010. Getting it 'better': The importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies.* **42**(3), pp.411-430.

Ryan, R.M. 1998. *Commentary: Human psychological needs and the issues of volition, control, and outcome focus.* New York, NY, US: Cambridge University Press.

Ryan, R.M. and Deci, E.L. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist.* **55**(1), pp.68-78.

Ryan, R.M. and Deci, E.L. 2009. Promoting self-determined school engagement: Motivation, learning, and well-being. In: Wenzel, K.R. and Wigfield, A. eds.

Handbook of motivation at school. New York, NY, US: Routledge/Taylor & Francis Group, pp.171-195.

Ryan, R.M. and Deci, E.L. 2017. *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* New York, NY, US: The Guilford Press.

Salili, F. and Hoosain, R. 2007. *Culture, motivation and learning: A multicultural perspective.* IAP.

Saruta, Y. 2010. *Ronri-teki-shikou ni motozuita kagaku-teki-hyougen-ryoku ni kansuru kenkyu*. PhD thesis, Hiroshima University.

Schiefele, U., Krapp, A. and Winteler, A. 1992. Interest as a predictor of academic achievement: A meta-analysis of research. *The role of interest in learning and development.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, pp.183-212.

Schoon, I. and Polek, E. 2011. Teenage career aspirations and adult career attainment: The role of gender, social background and general cognitive ability. *International Journal of Behavioral Development.* **35**(3), pp.210-217.

Shavelson, R.J., Hubner, J.J. and Stanton, G.C. 1976. Self-concept: Validation of construct interpretations. *Review of Educational Research.* **46**(3), pp.407-441.

Shen, C. and Pedulla, J.J. 2000. The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education: Principles, Policy & Practice.* **7**, pp.237-253.

Shen, C. and Tam, H.P. 2008. The paradoxical relationship between student achievement and self-perception: A cross-national analysis based on three waves of timss data. *Educational Research and Evaluation.* **14**(1), pp.87-100.

Shrout, P.E. and Bolger, N. 2002. Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods.* **7**(4), pp.422-445.

Simpkins, S.D., Davis-Kean, P.E. and Eccles, J.S. 2006. Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental psychology.* **42**(1), p70.

Simpkins, S.D., Fredricks, J.A. and Eccles, J.S. 2012. Charting the eccles' expectancy-value model from mothers' beliefs in childhood to youths' activities in adolescence. *Dev Psychol.* **48**(4), pp.1019-1032.

Singh, K., Granville, M. and Dika, S. 2002. Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research.* **95**(6), pp.323-332.

Sirin, S.R. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research.* **75**(3), pp.417-453.

Stapleton, L.M. 2013. Multilevel structural equation modeling with complex sample data. *Structural equation modeling: A second course, 2nd ed.* Charlotte, NC, US: IAP Information Age Publishing, pp.521-562.

Steenkamp, J.-B.E.M. and Baumgartner, H. 1998. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research.* **25**, pp.78-90.

Steinmayr, R. and Spinath, B. 2008. Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality.* **22**(3), pp.185-209.

Tavani, C.M. and Losh, S.C. 2003. Motivation, self-confidence, and expectations as predictors of the academic performances among our high school students. *Child Study Journal.* **33**(3), p141.

Toland, M.D. and De Ayala, R. 2005. A multilevel factor analysis of students' evaluations of teaching. *Educational and Psychological Measurement.* **65**(2), pp.272-296.

Tonks, S., Wigfield, A. and Eccles, J. 2018. Expectancy value theory in cross-cultural perspective: What have we learned in the last 15 years. In: Liem, G.A.D. and McInerney, D. eds. *Big theories revisited.* 2nd ed. Information Age Publisher.

Tóth-Király, I., Bõthe, B., Rigó, A. and Orosz, G. 2017. An illustration of the exploratory structural equation modeling (esem) framework on the passion scale. *Frontiers in Psychology.* **8**.

Trautwein, U., Lüdtke, O., Marsh, H.W., Köller, O. and Baumert, J. 2006. Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology.* **98**, pp.788-806.

Trautwein, U., Marsh, H.W., Nagengast, B., Lüdtke, O., Nagy, G. and Jonkmann, K. 2012. Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology.* **104**(3), pp.763-777.

Tripathy, J.P. 2013. Secondary data analysis: Ethical issues and challenges. *Iran J Public Health.* **42**(12), pp.1478-1479.

TTKB. 2018. *Matematik (1-8. Sınıflar) dersi öğretim programı [mathematics lesson (grades 1-8) curriculum].*

UCONN. No Date. *Guidance on secondary analysis of existing data sets.* [Online]. [Accessed 27.07.2023].

UoL. 2021. *University of leeds research ethics policy.* [Online]. [Accessed 27 February 2023]. Available from: https://ris.leeds.ac.uk/research-ethics-and-integrity/ethics-and-ethical-review/

Valentine, J.C., DuBois, D.L. and Cooper, H. 2004. The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist.* **39**(2), pp.111-133.

Van de Vijver, F.J. and Leung, K. 2000. Methodological issues in psychological research on culture. *Journal of cross-cultural psychology.* **31**(1), pp.33-51.

van Zyl, L.E. and Ten Klooster, P.M. 2021. Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for mplus. *Front Psychiatry.* **12**, p795672.

VanderWeele, T.J. 2016. Mediation analysis: A practitioner's guide. *Annu Rev Public Health.* **37**, pp.17-32.

Vansteenkiste, M., Timmermans, T., Lens, W., Soenens, B. and Van den Broeck, A. 2008. Does extrinsic goal framing enhance extrinsic goal-oriented individuals' learning and performance? An experimental test of the match perspective versus self-determination theory. *Journal of Educational Psychology.* **100**, pp.387-397.

Voyer, D. and Voyer, S.D. 2014. Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin.* **140**, pp.1174-1204.

Wang, M.-T. and Eccles, J.S. 2012. Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence.* **22**(1), pp.31-39.

Wang, M.T. and Degol, J. 2013. Motivational pathways to stem career choices: Using expectancy-value perspective to understand individual and gender differences in stem fields. *Dev Rev.* **33**(4).

Wang, T.L. and Berlin, D. 2010. Construction and validation of an instrument to measure taiwanese elementary students' attitudes toward their science class. *International Journal of Science Education.* **32**(18), pp.2413-2428.

Watt, H.M. 2004. Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade australian students. *Child development.* **75**(5), pp.1556-1574.

Watt, H.M.G., Shapka, J.D., Morris, Z.A., Durik, A.M., Keating, D.P. and Eccles, J.S. 2012. Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from australia, canada, and the united states. *Developmental Psychology.* **48**, pp.1594-1611.

White, K.R. 1982. The relation between socioeconomic status and academic achievement. *Psychological Bulletin.* **91**, pp.461-481.

Widaman, K.F. and Reise, S.P. 1997. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research.* Washington, DC, US: American Psychological Association, pp.281-324.

Wigfield, A. 1994. Expectancy-value theory of achievement motivation: A developmental perspective. *Educational psychology review.* **6**(1), pp.49-78.

Wigfield, A. and Eccles, J.S. 2000. Expectancy–value theory of achievement motivation. *Contemporary educational psychology.* **25**(1), pp.68-81.

Wigfield, A. and Eccles, J.S. 2002. The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence.

Wigfield, A., Eccles, J.S., Yoon, K.S., Harold, R.D., Arbreton, A.J.A., Freedman-Doan, C. and Blumenfeld, P.C. 1997. Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology.* **89**, pp.451-469.

Wigfield, A., Tonks, S. and Klauda, S.L. 2009. Expectancy-value theory. *Handbook of motivation at school.* New York, NY, US: Routledge/Taylor & Francis Group, pp.55-75.

Wigfield, A., Tonks, S.M. and Klauda, S.L. 2016. Expectancy-value theory. *Handbook of motivation at school.* Routledge.

Wilkins, J.L.M., Zembylas, M. and Travers, K.J. 2002. Investigating correlates of mathematics and science literacy in the final year of secondary school. In: Robitaille, D.F. and Beaton, A.E. eds. *Secondary analysis of the timss data.* Dordrecht: Springer Netherlands, pp.291-316.

Xiao, Y., Liu, H. and Hau, K.-T. 2019. A comparison of cfa, esem, and bsem in test structure analysis. *Structural Equation Modeling: A Multidisciplinary Journal.* **26**(5), pp.665-677.

Yang, Y. and Green, S.B. 2011. Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment.* **29**(4), pp.377-392.

Ye, F. and Wallace, T.L. 2014. Psychological sense of school membership scale:Method effects associated with negatively worded items. *Journal of Psychoeducational Assessment.* **32**(3), pp.202-215.

Yin, L. and Fishbein, B. 2019. Creating and interpreting the timss 2019 context questionnaire scales. In: Martin, M.O., et al. eds. *Methods and procedures: Timss 2019 technical report* Boston College, pp.16.11-16.331.

Yoshino, A. 2012. The relationship between self-concept and achievement in timss 2007: A comparison between american and japanese students. *International Review of*

Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education. **58**(2), pp.199-219.

Zhu, Y. and Leung, F.K.S. 2011. Mathematics achievement of mainland immigrant students in hong kong. *Asia Pacific Journal of Education.* **31**(4), pp.471-485.

Appendix 1

The Secretariat
University of Leeds
Leeds, LS2 9JT Tel: 0113 343 4873
Email: ResearchEthics@leeds.ac.uk

UNIVERSITY OF LEEDS

Tevfik Karabiyik
School of Social Sciences and Law
University of Leeds
Leeds, LS2 9JT

ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee
University of Leeds

14 August 2023

Dear Tevfik,

Title of study: **Factors Associated with Mathematics Attainment: A Comparative Psychometric Study of Türkiye, England and Japan**
Ethics reference: **LTEDUC-101**

I am pleased to inform you that the above application for light touch ethical review has been reviewed by a representative of the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee and I can confirm a favourable ethical opinion as of the date of this letter. The following documentation was considered:

Document	Version	Date
LTEDUC-101 LightTouchEthicsForm tevfik_ (002).doc	1	18/09/2018

Please notify the committee if you intend to make any amendments to the original research as submitted at date of this approval, including changes to recruitment methodology. All changes must receive ethical approval prior to implementation. The amendment form is available at http://ris.leeds.ac.uk/EthicsAmendment.

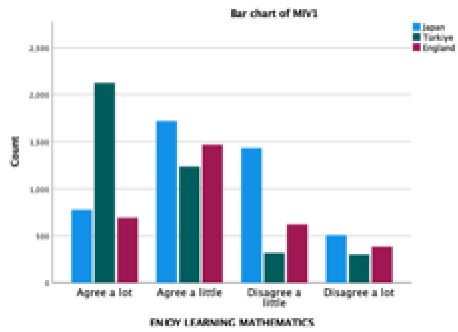
Please note: You are expected to keep a record of all your approved documentation, as well as other documents relating to the study. You will be given a two week notice period if your project is to be audited, there is a checklist listing examples of documents to be kept which is available at http://ris.leeds.ac.uk/EthicsAudits.

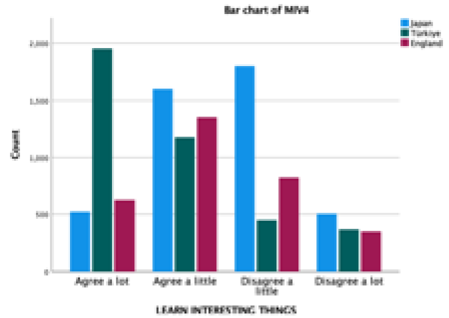
We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to ResearchEthics@leeds.ac.uk.

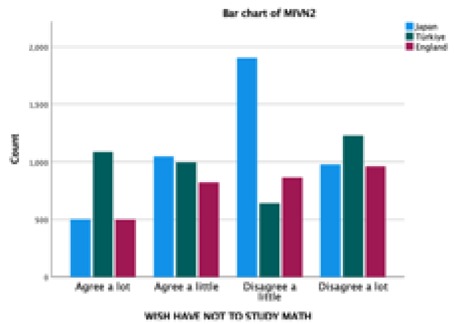
Yours sincerely
Jennifer Blaikie
Senior Research Ethics Administrator, the Secretariat
On behalf of Dr Kahryn Hughes, Chair, AREA Faculty Research Ethics Committee

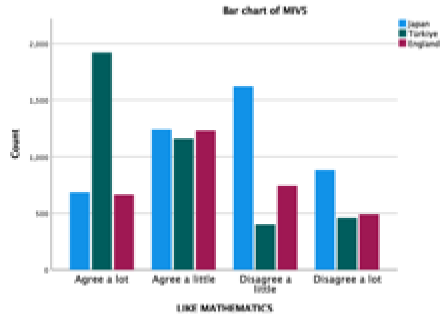
<u>Appendix 2</u>

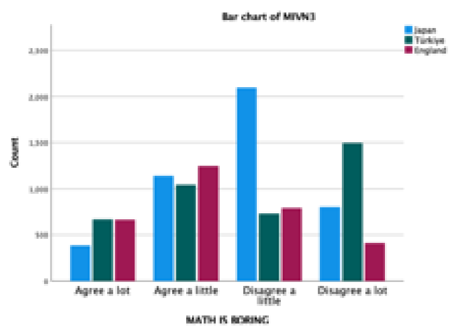
Bar chart of the responses to the Mathematics Intrinsic Motivation (MIV) scale

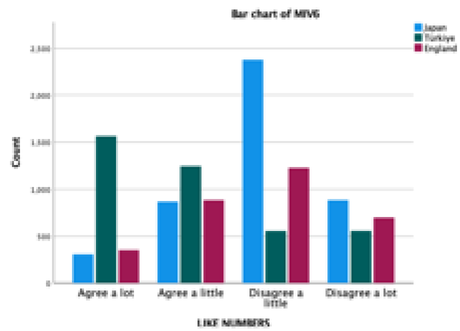


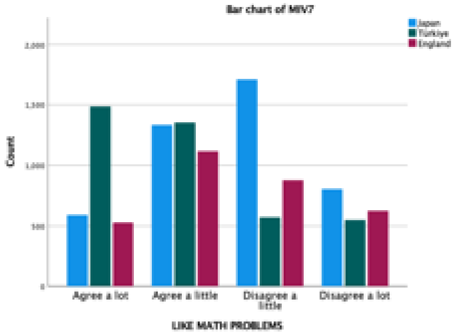


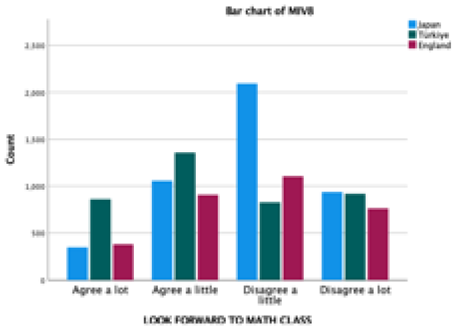


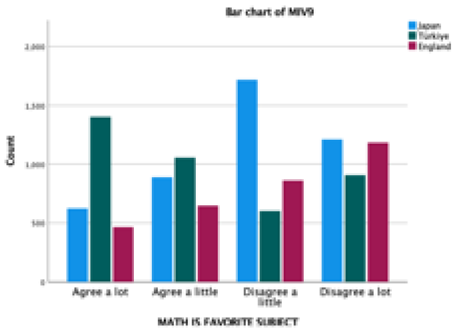




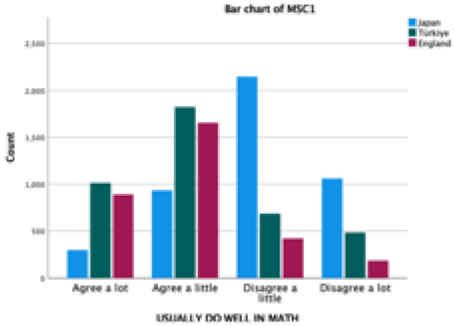


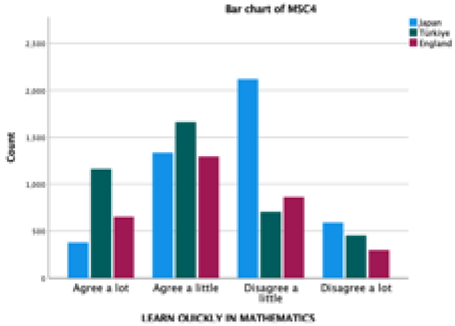

Bar chart of MIV7
Japan
Türkiye
England
Count
Agree a lot
Agree a little
Disagree a little
Disagree a lot
LIKE MATH PROBLEMS

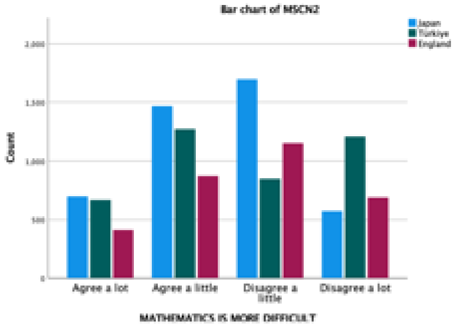

Bar chart of MIV8
Japan
Türkiye
England
Count
Agree a lot
Agree a little
Disagree a little
Disagree a lot
LOOK FORWARD TO MATH CLASS

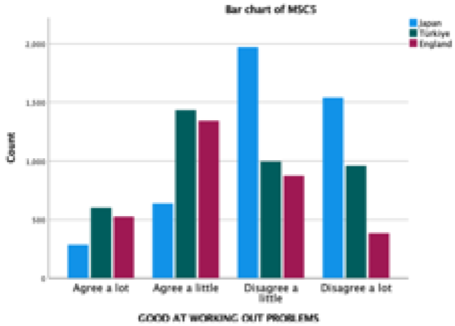

Bar chart of MIV9
Japan
Türkiye
England
Count
Agree a lot
Agree a little
Disagree a little
Disagree a lot
MATH IS FAVORITE SUBJECT

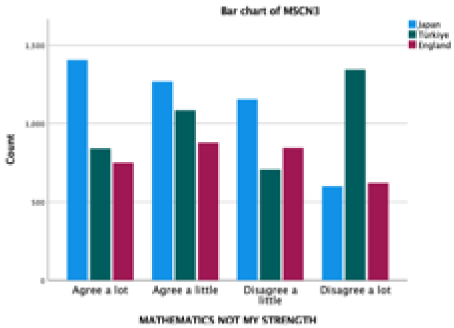
Bar chart of the responses to the Mathematics Self-Concept (MSC) scale

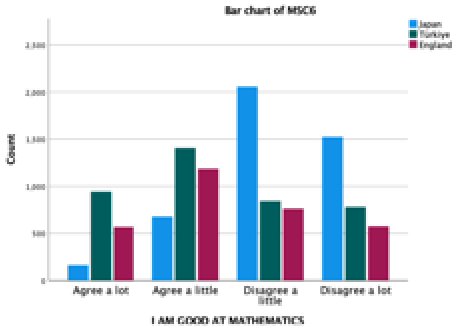


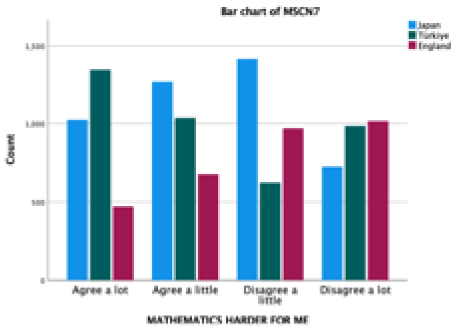


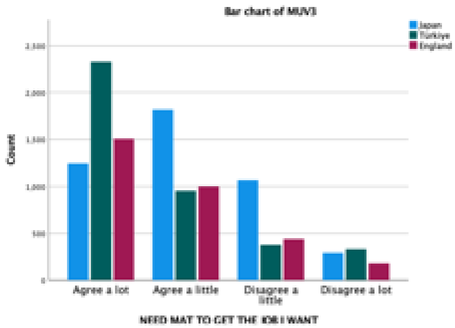




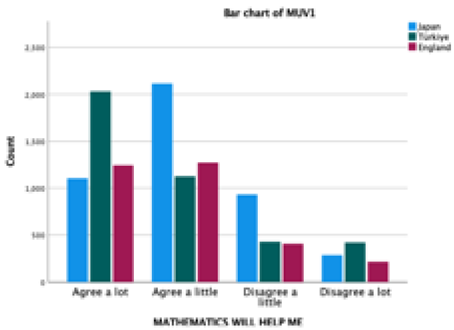


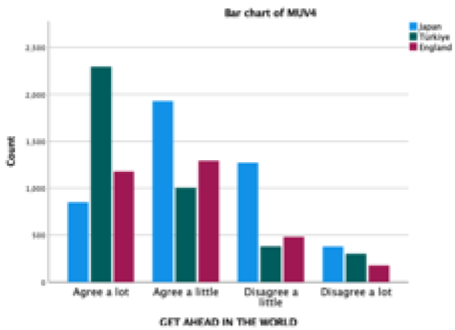


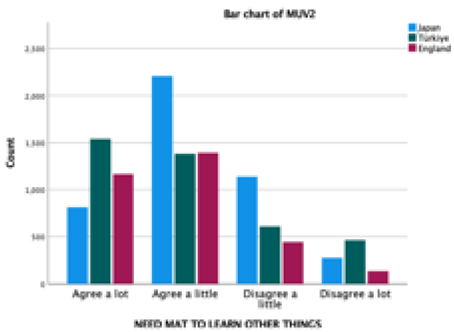


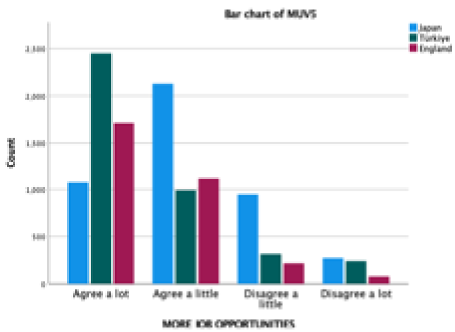


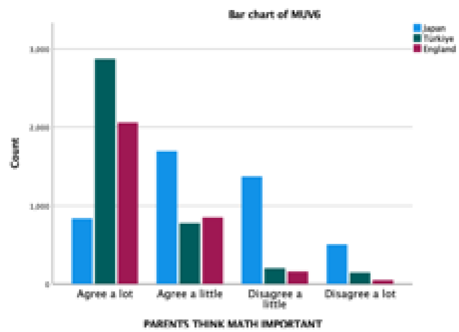
Bar chart of the responses to the Mathematics Utility Value (MUV) scale

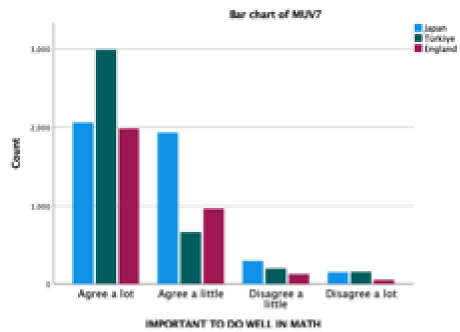




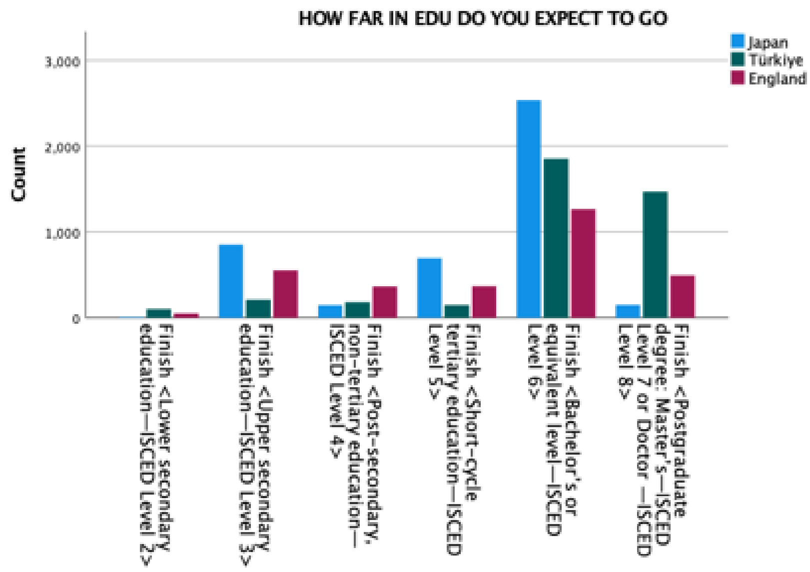




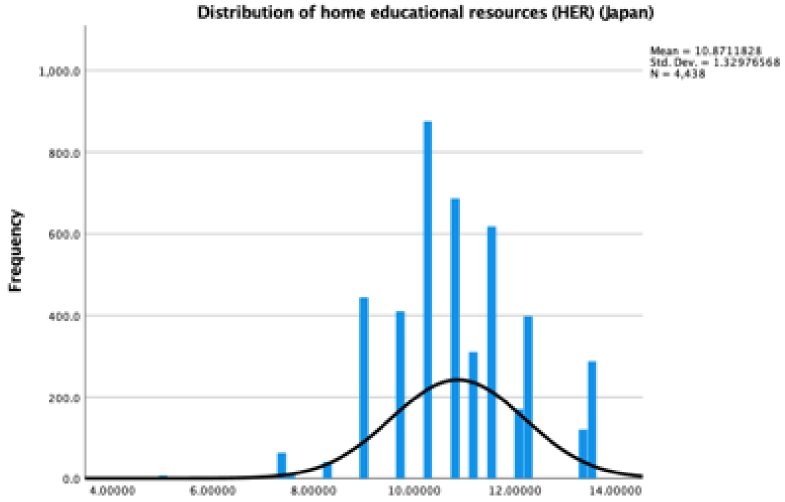


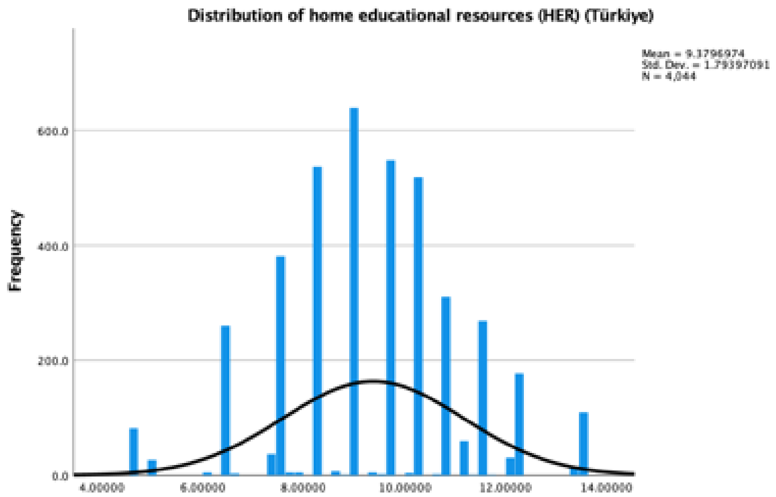


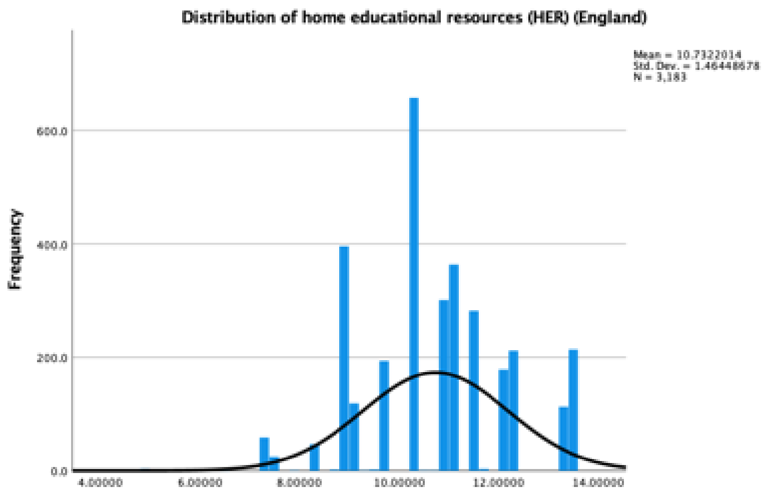
Bar chart of the responses to the Educational Aspiration (EDU ASP)



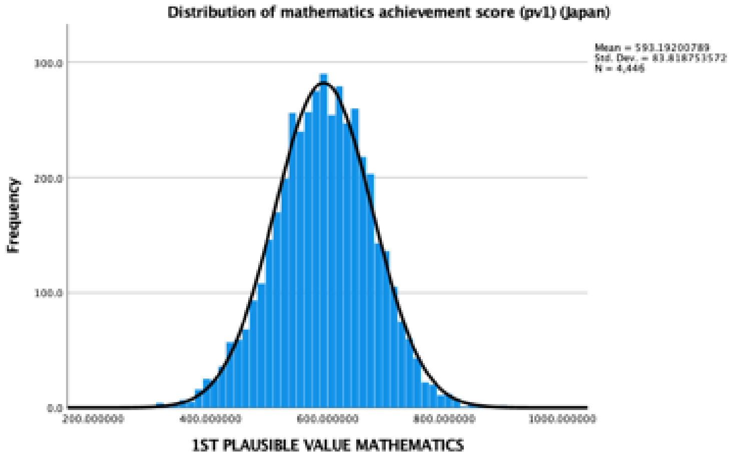
Histogram graph of home educational resources (HER)

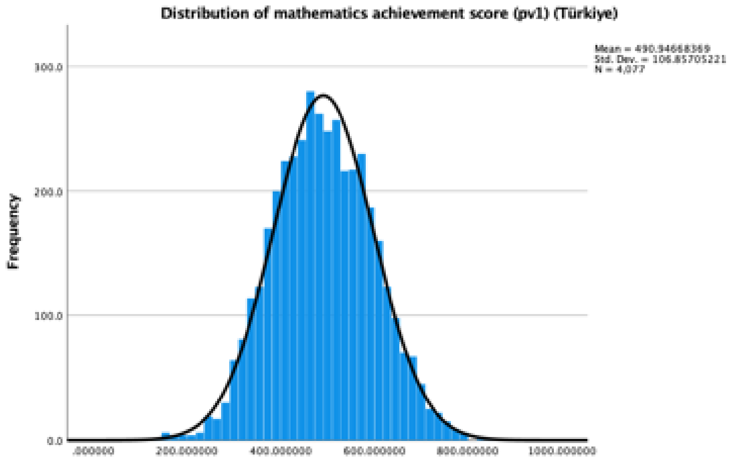


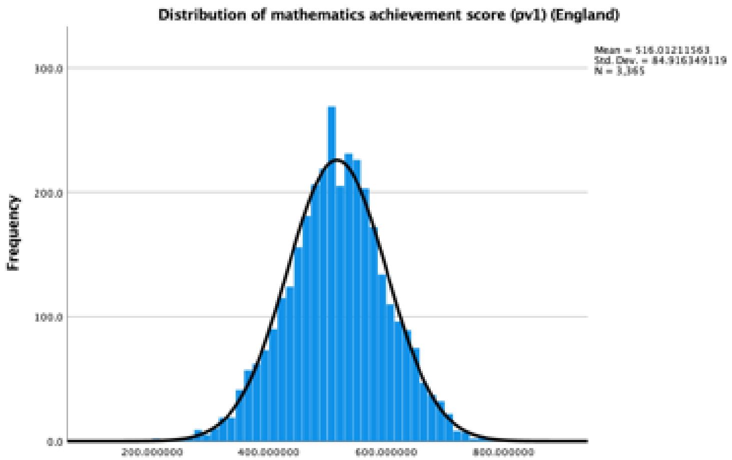




Histogram graph of mathematics achievement (pv1) score







Appendix 3

The appendix includes the Mplus syntaxes corresponding to the analysis presented in chapter 4.

Model 1

```
TITLE: Total CFA (Model 1);
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      MISSING ARE ALL(-999);

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR = MLR;

MODEL:      MIV by MIV1* MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
MIV9;
        MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;
        MUV by MUV1* MUV2 MUV3 MUV4 MUV5 MUV6 MUV7;
        MIV@1; MSC@1; MUV@1;

OUTPUT:   SAMPSTAT MODINDICES STDYX;
```

Model 2

```
TITLE: Total CFA with correlated uniqueness;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      MISSING ARE ALL(-999);

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;

        MODEL:     MIV by MIV1* MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7
                   MIV8 MIV9;
        MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;
        MUV by MUV1* MUV2 MUV3 MUV4 MUV5 MUV6 MUV7;
        MIV@1; MSC@1; MUV@1;

        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

OUTPUT:   SAMPSTAT MODINDICES STDYX;
```

Model 3

```
TITLE: ESEM;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
    NAMES ARE ITSEX HER_SCL EDU_ASP
    MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
    MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
    MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
    MAT_ACH IDCLASS HOUWGT IDCNTRY;

    usevariables are
    MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
    MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
    MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 CNTCLS;

    CLUSTER IS CNTCLS;
    WEIGHT IS HOUWGT;
    MISSING ARE ALL(-999);

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR = MLR;
ROTATION = TARGET;
! The factors are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but
targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)

 MODEL:     MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
            MIV9
            MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0
            MUV7~0 (*1);
            MSC by MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
            MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
     MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0 MIV8~0
     MIV9~0 (*1) ;
     OUTPUT:   SAMPSTAT SVALUES TECH4 STDYX;
```

Model 4

```
TITLE: ESEM with CU;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      MISSING ARE ALL(-999);

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR = MLR;
ROTATION = TARGET;
! The factors are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but
targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)

 MODEL:      MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
             MIV9
             MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0
             MUV7~0 (*1);
             MSC by MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
             MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
        MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0 MIV8~0
        MIV9~0 (*1) ;

                     MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
                     MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

OUTPUT:    SAMPSTAT SVALUES TECH4 STDYX;
```

Model 5

```
TITLE: CFA configural invariance;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);

ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        model = configural;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;

MODEL:       MIV by MIV1* MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
MIV9;
        MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;
        MUV by MUV1* MUV2 MUV3 MUV4 MUV5 MUV6 MUV7;
        MIV@1; MSC@1; MUV@1;

        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

OUTPUT:   SAMPSTAT STDYX;
```

Model 6

```
TITLE: CFA metric invariance;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);


ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        model = METRIC;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;

MODEL:       MIV by MIV1* MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
                   MIV9;
             MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;
             MUV by MUV1* MUV2 MUV3 MUV4 MUV5 MUV6 MUV7;
             MIV@1; MSC@1; MUV@1;

             MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
             MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

OUTPUT:   SAMPSTAT STDYX;
```

Model 7

```
TITLE: ESEM configural inv;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);

ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        ROTATION = TARGET;
        model = configural;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
! The factors  are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)
MODEL:    MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8     MIV9
          MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0
          MUV7~0 (*1);
          MSC by  MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
          MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
          MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0
          MIV8~0 MIV9~0 (*1);

          MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
          MIVN2 MIVN3 MSCN2 MSCN3 MSCN7;

OUTPUT:   SAMPSTAT STDYX;
```

Model 8

```
TITLE: ESEM metric inv;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);
ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        ROTATION = TARGET;
        model = metric;
DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
! The factors are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)

MODEL:       MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
                      MIV9  MUV1~0  MUV2~0  MUV3~0  MUV4~0  MUV5~0
                      MUV6~0 MUV7~0 (*1);

             MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
             MSC@1;
             MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
                       MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0
                       MIV7~0 MIV8~0 MIV9~0 (*1);

MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

OUTPUT:    SAMPSTAT STDYX;
```

Model 9

```
TITLE: ESEM scalar inv;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);
ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        ROTATION = TARGET;
        model = scalar;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
! The factors  are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)

 MODEL:      MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
                     MIV9 MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0
                     MUV6~0 MUV7~0 (*1);

             MSC by MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;

             MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
                      MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0
                      MIV7~0 MIV8~0 MIV9~0 (*1);

        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7;
```

Model 10

```
TITLE: ESEM partial scalar;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      CNTCLS;

      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);
ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        ROTATION = TARGET;
        model = scalar;
DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
! The factors are defined with main loadings from their respective items
! In addition to these main loadings, all other cross-loadings are estimated but targeted
! to be as close to 0 as possible (~0)
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1)

 MODEL:       MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
                        MIV9 MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0
                        MUV6~0 MUV7~0 (*1);
              MSC by MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
              MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
                        MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0
                        MIV7~0 MIV8~0 MIV9~0 (*1);
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

! freely estimate intercepts of below selected items
     [MIVN2 MIVN3 MIV6 MIV9]; [MSC1 MSC5 MSCN7]; [muv2 MUV6];
```

Model 11

```
TITLE: ESEM MIMIC model;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
       NAMES ARE ITSEX HER_SCL EDU_ASP
       MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
       MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
       MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
       MAT_ACH IDCLASS HOUWGT IDCNTRY;

       usevariables are
      ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH CNTCLS;

       CLUSTER IS CNTCLS;
       WEIGHT IS HOUWGT;
       GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
       MISSING ARE ALL(-999);
       ANALYSIS: TYPE IS COMPLEX;
       ESTIMATOR = MLR;
       ROTATION = TARGET;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;

 MODEL:       MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8
                     MIV9  MUV1~0  MUV2~0  MUV3~0  MUV4~0  MUV5~0
                     MUV6~0 MUV7~0 (*1);
              MSC by MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7;
              MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
                      MIV1~0  MIVN2~0  MIVN3~0  MIV4~0  MIV5~0  MIV6~0
                      MIV7~0 MIV8~0 MIV9~0 (*1);

         MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
         MIVN2 MIVN3 MSCN2 MSCN3 MSCN7;

        [MIVN2 MIVN3 MIV6 MIV9]; [MSC1 MSC5 MSCN7]; [muv2 MUV6];
         MIV MSC MUV, ITSEX HER_SCL EDU_ASP MAT_ACH with
         MIV MSC MUV, ITSEX HER_SCL EDU_ASP MAT_ACH;

OUTPUT:    SAMPSTAT SVALUES TECH4 STDYX;
```

Appendix 4

The appendix includes the Mplus syntaxes corresponding to the analysis presented in chapter 5.

Model 12

```
TITLE: ESEM task values;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
       NAMES ARE ITSEX HER_SCL EDU_ASP
       MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
       MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
       MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
       MAT_ACH IDCLASS HOUWGT IDCNTRY;

      usevariables are
      EDU_ASP MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH CNTCLS;

       CLUSTER IS CNTCLS;
       WEIGHT IS HOUWGT;
       GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
       MISSING ARE ALL(-999);
       ANALYSIS: TYPE IS COMPLEX;
       ESTIMATOR = MLR;
       ROTATION = TARGET;

DEFINE:   CNTCLS = (IDCNTRY*1000000) + IDCLASS;
MODEL:MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
          MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0 MUV7~0 (*1);
          MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MIV1~0
          MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0 MIV8~0
          MIV9~0 (*1);

MIVN2 MIVN3 WITH MIVN2 MIVN3 ;
 [MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9];
 [MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7];
 [MIV@0]; [MUV@0];
EDU_ASP ON MIV MUV;
MAT_ACH ON MIV MUV;

OUTPUT: SAMPSTAT Svalues TECH4 STDYX;
```

Model 13

```
TITLE: ESEM task value+expectancy;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
       NAMES ARE ITSEX HER_SCL EDU_ASP
       MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
       MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
       MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
       MAT_ACH IDCLASS HOUWGT IDCNTRY;

     usevariables are
     EDU_ASP MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
     MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
     MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH CNTCLS;

       CLUSTER IS CNTCLS;
       WEIGHT IS HOUWGT;
       GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
       MISSING ARE ALL(-999);

ANALYSIS: TYPE IS COMPLEX;
                ESTIMATOR = MLR;
                ROTATION = TARGET;
DEFINE:      CNTCLS = (IDCNTRY*1000000)+IDCLASS;
MODEL: MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
         MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0 MUV7~0 (*1);
         MSC by  MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
         MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
         MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0 MIV8~0
         MIV9~0 (*1);

MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;  MSC@1;
 [MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9];
 [MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7];
 [MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7];
 [MIV@0]; [MSC@0]; [MUV@0];
 EDU_ASP ON MIV MUV MSC ;
 MAT_ACH ON MIV MUV MSC ;
OUTPUT: SAMPSTAT Svalues TECH4 STDYX;
```

Model 14

```
TITLE: ESEM mscxmiv interaction;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
        NAMES ARE ITSEX HER_SCL EDU_ASP
        MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
        MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
        MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
        MAT_ACH IDCLASS HOUWGT IDCNTRY;
        usevariables are
        EDU_ASP MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
        MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
        MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH
        CNTCLS
        S1XI7 S2XI4 S3XI5 S4XI9 S6XI1 S7XI3 S8XI6;

        CLUSTER IS CNTCLS;
        WEIGHT IS HOUWGT;
        GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
        MISSING ARE ALL(-999);

ANALYSIS: TYPE IS COMPLEX;
          ESTIMATOR = MLR;
          ROTATION = TARGET ;

DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
! here we centered the items to create interaction term
         if (idcntry eq 392) then cmiv1 = miv1-2.624;
         if (idcntry eq 392) then cmiv3 = mivn3-2.748;
         if (idcntry eq 392) then cmiv4 = miv4-2.481;
         if (idcntry eq 392) then cmiv5 = miv5-2.388;
         if (idcntry eq 392) then cmiv6 = miv6-2.129;
         if (idcntry eq 392) then cmiv7 = miv7-2.624;
         if (idcntry eq 392) then cmiv9 = miv9-2.206;
         if (idcntry eq 392) then cmsc1 = msc1-2.109;
         if (idcntry eq 392) then cmsc2 = mscn2-2.485;
         if (idcntry eq 392) then cmsc3 = mscn3-2.214;
         if (idcntry eq 392) then cmsc4 = msc4-2.338;
         if (idcntry eq 392) then cmsc5 = msc5-1.928;
         if (idcntry eq 392) then cmsc6 = msc6-1.884;
         if (idcntry eq 392) then cmsc7 = mscn7-2.419;
         if (idcntry eq 792) then cmiv1 = miv1-3.300;
         if (idcntry eq 792) then cmiv3 = mivn3-2.764;
         if (idcntry eq 792) then cmiv4 = miv4-3.195;
         if (idcntry eq 792) then cmiv5 = miv5-3.148;
         if (idcntry eq 792) then cmiv6 = miv6-2.953;
         if (idcntry eq 792) then cmiv7 = miv7-2.951;
```

```
if (idcntry eq 792) then cmiv9 = miv9-2.740;
if (idcntry eq 792) then cmsc1 = msc1-2.845;
if (idcntry eq 792) then cmsc2 = mscn2-2.662;
if (idcntry eq 792) then cmsc3 = mscn3-2.655;
if (idcntry eq 792) then cmsc4 = msc4-2.893;
if (idcntry eq 792) then cmsc5 = msc5-2.421;
if (idcntry eq 792) then cmsc6 = msc6-2.652;
if (idcntry eq 792) then cmsc7 = mscn7-2.324;
if (idcntry eq 926) then cmiv1 = miv1-2.798;
if (idcntry eq 926) then cmiv3 = mivn3-2.320;
if (idcntry eq 926) then cmiv4 = miv4-2.731;
if (idcntry eq 926) then cmiv5 = miv5-2.681;
if (idcntry eq 926) then cmiv6 = miv6-2.284;
if (idcntry eq 926) then cmiv7 = miv7-2.502;
if (idcntry eq 926) then cmiv9 = miv9-2.144;
if (idcntry eq 926) then cmsc1 = msc1-3.027;
if (idcntry eq 926) then cmsc2 = mscn2-2.662;
if (idcntry eq 926) then cmsc3 = mscn3-2.417;
if (idcntry eq 926) then cmsc4 = msc4-2.739;
if (idcntry eq 926) then cmsc5 = msc5-2.636;
if (idcntry eq 926) then cmsc6 = msc6-2.581;
if (idcntry eq 926) then cmsc7 = mscn7-2.797;
S1XI7 = CMSC1*CMIV7;
S2XI4 = CMSC2*CMIV4;
S3XI5 = CMSC3*CMIV5;
S4XI9 = CMSC4*CMIV9;
S5XI1 = CMSC5*CMIV1;
S6XI3 = CMSC6*CMIV3;
S7XI6 = CMSC7*CMIV6;
MODEL: MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
           MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0 MUV7~0 (*1);
           MSC by MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
           MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
           MIV1~0 MIVN2~0 MIVN3~0 MIV4~0 MIV5~0 MIV6~0 MIV7~0
           MIV8~0 MIV9~0 (*1);
           SCXIV BY S1XI7* S2XI4 S3XI5 S4XI9 S5XI1 S6XI3 S7XI6;
           MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
           MIVN2 MIVN3 MSCN2 MSCN3 MSCN7;
miv with msc (1);
miv with muv;
muv with msc;
MSC@1; SCXIV@1;
 [MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9];
 [MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7];
 [MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7];
 [S1XI7 S2XI4 S3XI5 S4XI9 S5XI1 S6XI3 S7XI6];
 [MIV@0]; [MSC@0]; [MUV@0]; [SCXIV](1);
EDU_ASP ON MIV MSC MUV SCXIV;
MAT_ACH ON MIV MSC MUV SCXIV;
OUTPUT:    SAMPSTAT Svalues TECH4 STDYX;
```

Model 15

```
TITLE: ESEM mscxmuv interaction;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplus
results/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
        NAMES ARE ITSEX HER_SCL EDU_ASP
        MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
        MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
        MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
        MAT_ACH IDCLASS HOUWGT IDCNTRY;
       usevariables are
        EDU_ASP MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
        MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
        MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH CNTCLS
        S1XU2 S2XU1 S3XU3 S4XU6 S5XU5 S6XU4 S7XU7;
        CLUSTER IS CNTCLS;
        WEIGHT IS HOUWGT;
        GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
        MISSING ARE ALL(-999);
ANALYSIS: TYPE IS COMPLEX;
          ESTIMATOR = MLR;
          ROTATION = TARGET;
DEFINE:   CNTCLS = (IDCNTRY*1000000)+IDCLASS;
         if (idcntry eq 392) then cmsc1 = msc1-2.109;
         if (idcntry eq 392) then cmsc2 = mscn2-2.485;
         if (idcntry eq 392) then cmsc3 = mscn3-2.214;
         if (idcntry eq 392) then cmsc4 = msc4-2.338;
         if (idcntry eq 392) then cmsc5 = msc5-1.928;
         if (idcntry eq 392) then cmsc6 = msc6-1.884;
         if (idcntry eq 392) then cmsc7 = mscn7-2.419;
         if (idcntry eq 392) then cmuv1 = muv1-2.910;
         if (idcntry eq 392) then cmuv2 = muv2-2.801;
         if (idcntry eq 392) then cmuv3 = muv3-2.913;
         if (idcntry eq 392) then cmuv4 = muv4-2.736;
         if (idcntry eq 392) then cmuv5 = muv5-2.906;
         if (idcntry eq 392) then cmuv6 = muv6-2.648;
         if (idcntry eq 392) then cmuv7 = muv7-3.328;
         if (idcntry eq 792) then cmsc1 = msc1-2.845;
         if (idcntry eq 792) then cmsc2 = mscn2-2.662;
         if (idcntry eq 792) then cmsc3 = mscn3-2.655;
         if (idcntry eq 792) then cmsc4 = msc4-2.893;
         if (idcntry eq 792) then cmsc5 = msc5-2.421;
         if (idcntry eq 792) then cmsc6 = msc6-2.652;
         if (idcntry eq 792) then cmsc7 = mscn7-2.324;
         if (idcntry eq 792) then cmuv1 = muv1-3.194;
         if (idcntry eq 792) then cmuv2 = muv2-2.997;
         if (idcntry eq 792) then cmuv3 = muv3-3.322;
         if (idcntry eq 792) then cmuv4 = muv4-2.331;
```

```
        if (idcntry eq 792) then cmuv5 = muv5-3.415;
        if (idcntry eq 792) then cmuv6 = muv6-3.604;
        if (idcntry eq 792) then cmuv7 = muv7-3.620;
        if (idcntry eq 926) then cmsc1 = msc1-3.027;
        if (idcntry eq 926) then cmsc2 = mscn2-2.662;
        if (idcntry eq 926) then cmsc3 = mscn3-2.417;
        if (idcntry eq 926) then cmsc4 = msc4-2.739;
        if (idcntry eq 926) then cmsc5 = msc5-2.636;
        if (idcntry eq 926) then cmsc6 = msc6-2.581;
        if (idcntry eq 926) then cmsc7 = mscn7-2.797;
        if (idcntry eq 926) then cmuv1 = muv1-3.133;
        if (idcntry eq 926) then cmuv2 = muv2-3.148;
        if (idcntry eq 926) then cmuv3 = muv3-3.226;
        if (idcntry eq 926) then cmuv4 = muv4-3.111;
        if (idcntry eq 926) then cmuv5 = muv5-3.420;
        if (idcntry eq 926) then cmuv6 = muv6-3.565;
        if (idcntry eq 926) then cmuv7 = muv7-3.557;
        S1XU2 = CMSC1*CMUV2;
        S2XU1 = CMSC2*CMUV1;
        S3XU3 = CMSC3*CMUV3;
        S4XU6 = CMSC4*CMUV6;
        S5XU5 = CMSC5*CMUV5;
        S6XU4 = CMSC6*CMUV4;
        S7XU7 = CMSC7*CMUV7;
MODEL:MIV by MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
          MUV1~0 MUV2~0 MUV3~0 MUV4~0 MUV5~0 MUV6~0 MUV7~0 (*1);
          MSC by  MSC1* MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7 ;
          MUV by MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MIV1~0
          MIVN2~0   MIVN3~0   MIV4~0   MIV5~0   MIV6~0   MIV7~0   MIV8~0
        MIV9~0 (*1);
        SCXUV BY S1XU2* S2XU1 S3XU3 S4XU6 S5XU5 S6XU4 S7XU7;
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
        MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;
       miv with muv;
       muv with msc;
       muv with msc (2);
       EDU_ASP WITH MAT_ACH;
       MSC@1; SCXUV@1;
      [MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9];
      [MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7];
      [MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7];
      [S1XU2 S2XU1 S3XU3 S4XU6 S5XU5 S6XU4 S7XU7];
      [MIV@0]; [MSC@0]; [MUV@0]; [SCXUV](2);
      EDU_ASP ON MIV MSC MUV SCXUV;
      MAT_ACH ON MIV MSC MUV SCXUV;

OUTPUT:    SAMPSTAT Svalues TECH4 STDYX;
```

Appendix 5

The appendix includes the Mplus syntax corresponding to the analysis presented in chapter 6.

Model 16

```
TITLE: ESEM mediation model;
DATA: FILE IS
"/Users/tevfikcankarabiyik/Desktop/tez/Mplusresults/imputation/timss19pv_list.dat";
TYPE=IMPUTATION;
VARIABLE:
      NAMES ARE ITSEX HER_SCL EDU_ASP
      MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9
      MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7
      MAT_ACH IDCLASS HOUWGT IDCNTRY;
      usevariables are
      ITSEX HER_SCL EDU_ASP MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7
      MIV8 MIV9 MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7
      MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7 MAT_ACH
      CNTCLS S1XI7 S2XI4 S3XI5 S4XI9 S5XI1 S6XI3 S7XI6
      S1XU2 S2XU1 S3XU3 S4XU6 S5XU5 S6XU4 S7XU7;
      CLUSTER IS CNTCLS;
      WEIGHT IS HOUWGT;
      GROUPING IS IDCNTRY ( 392=Japan 792=Türkiye 926=England);
      MISSING ARE ALL(-999);
ANALYSIS: TYPE IS COMPLEX;
        ESTIMATOR = MLR;
        H1ITERATIONS=20000;
        ITERATIONS=100000;
DEFINE:    CNTCLS = (IDCNTRY*1000000)+IDCLASS;
       if (idcntry eq 392) then cmiv1 = miv1-2.624;
       if (idcntry eq 392) then cmiv3 = mivn3-2.748;
       if (idcntry eq 392) then cmiv4 = miv4-2.481;
       if (idcntry eq 392) then cmiv5 = miv5-2.388;
       if (idcntry eq 392) then cmiv6 = miv6-2.129;
       if (idcntry eq 392) then cmiv7 = miv7-2.624;
       if (idcntry eq 392) then cmiv9 = miv9-2.206;
       if (idcntry eq 392) then cmsc1 = msc1-2.109;
       if (idcntry eq 392) then cmsc2 = mscn2-2.485;
       if (idcntry eq 392) then cmsc3 = mscn3-2.214;
       if (idcntry eq 392) then cmsc4 = msc4-2.338;
       if (idcntry eq 392) then cmsc5 = msc5-1.928;
       if (idcntry eq 392) then cmsc6 = msc6-1.884;
       if (idcntry eq 392) then cmsc7 = mscn7-2.419;
       if (idcntry eq 392) then cmuv1 = muv1-2.910;
       if (idcntry eq 392) then cmuv2 = muv2-2.801;
```

```
if (idcntry eq 392) then cmuv3 = muv3-2.913;
if (idcntry eq 392) then cmuv4 = muv4-2.736;
if (idcntry eq 392) then cmuv5 = muv5-2.906;
if (idcntry eq 392) then cmuv6 = muv6-2.648;
if (idcntry eq 392) then cmuv7 = muv7-3.328;
if (idcntry eq 792) then cmiv1 = miv1-3.300;
if (idcntry eq 792) then cmiv3 = mivn3-2.764;
if (idcntry eq 792) then cmiv4 = miv4-3.195;
if (idcntry eq 792) then cmiv5 = miv5-3.148;
if (idcntry eq 792) then cmiv6 = miv6-2.953;
if (idcntry eq 792) then cmiv7 = miv7-2.951;
if (idcntry eq 792) then cmiv9 = miv9-2.740;
if (idcntry eq 792) then cmsc1 = msc1-2.845;
if (idcntry eq 792) then cmsc2 = mscn2-2.662;
if (idcntry eq 792) then cmsc3 = mscn3-2.655;
if (idcntry eq 792) then cmsc4 = msc4-2.893;
if (idcntry eq 792) then cmsc5 = msc5-2.421;
if (idcntry eq 792) then cmsc6 = msc6-2.652;
if (idcntry eq 792) then cmsc7 = mscn7-2.324;
if (idcntry eq 792) then cmuv1 = muv1-3.194;
if (idcntry eq 792) then cmuv2 = muv2-2.997;
if (idcntry eq 792) then cmuv3 = muv3-3.322;
if (idcntry eq 792) then cmuv4 = muv4-2.331;
if (idcntry eq 792) then cmuv5 = muv5-3.415;
if (idcntry eq 792) then cmuv6 = muv6-3.604;
if (idcntry eq 792) then cmuv7 = muv7-3.620;
if (idcntry eq 926) then cmiv1 = miv1-2.798;
if (idcntry eq 926) then cmiv3 = mivn3-2.320;
if (idcntry eq 926) then cmiv4 = miv4-2.731;
if (idcntry eq 926) then cmiv5 = miv5-2.681;
if (idcntry eq 926) then cmiv6 = miv6-2.284;
if (idcntry eq 926) then cmiv7 = miv7-2.502;
if (idcntry eq 926) then cmiv9 = miv9-2.144;
if (idcntry eq 926) then cmsc1 = msc1-3.027;
if (idcntry eq 926) then cmsc2 = mscn2-2.662;
if (idcntry eq 926) then cmsc3 = mscn3-2.417;
if (idcntry eq 926) then cmsc4 = msc4-2.739;
if (idcntry eq 926) then cmsc5 = msc5-2.636;
if (idcntry eq 926) then cmsc6 = msc6-2.581;
if (idcntry eq 926) then cmsc7 = mscn7-2.797;
if (idcntry eq 926) then cmuv1 = muv1-3.133;
if (idcntry eq 926) then cmuv2 = muv2-3.148;
if (idcntry eq 926) then cmuv3 = muv3-3.226;
if (idcntry eq 926) then cmuv4 = muv4-3.111;
if (idcntry eq 926) then cmuv5 = muv5-3.420;
if (idcntry eq 926) then cmuv6 = muv6-3.565;
if (idcntry eq 926) then cmuv7 = muv7-3.557;

S1XI7 = CMSC1*CMIV7;
S2XI4 = CMSC2*CMIV4;
```

```
        S3XI5 = CMSC3*CMIV5;
        S4XI9 = CMSC4*CMIV9;
        S5XI1 = CMSC5*CMIV1;
        S6XI3 = CMSC6*CMIV3;
        S7XI6 = CMSC7*CMIV6;
        S1XU2 = CMSC1*CMUV2;
        S2XU1 = CMSC2*CMUV1;
        S3XU3 = CMSC3*CMUV3;
        S4XU6 = CMSC4*CMUV6;
        S5XU5 = CMSC5*CMUV5;
        S6XU4 = CMSC6*CMUV4;
        S7XU7 = CMSC7*CMUV7;

MODEL: miv BY miv1*0.765 ;
    miv BY mivn2*0.514 ;
    miv BY mivn3*0.608 ;
    miv BY miv4*0.569;
    miv BY miv5@0.909 ;
    miv BY miv6*0.676 ;
    miv BY miv7*0.819 ;
    miv BY miv8*0.739 ;
    miv BY miv9*0.933 ;
    miv BY muv1*0.157;
    miv BY muv2*-0.098;
    miv BY muv4*-0.010;
    miv BY muv6*0.001;
    miv BY muv7@-0.074;
    miv BY muv8*-0.076;
    miv BY muv9*-0.023;
    msc BY msc1*0.682 ;
    msc BY mscn2*0.506 ;
    msc BY mscn3*0.732 ;
    msc BY msc4*0.673 ;
    msc BY msc5@0.721 ;
    msc BY msc6*0.517 ;
    msc BY mscn7*0.625 ;
    muv BY muv1*0.485 ;
    muv BY muv2*0.491 ;
    muv BY muv3*0.630 ;
    muv BY muv4@0.729 ;
    muv BY muv5*0.696 ;
    muv BY muv6*0.456 ;
    muv BY muv7*0.524 ;
    muv BY miv1*-0.023;
    muv BY mivn2*0.144;
    muv BY mivn3*0.042;
    muv BY miv4*0.098;
    muv BY miv5*-0.050;
    muv BY miv6*0.016;
    muv BY miv7*-0.020;
```

```
muv BY miv8@-0.002;
muv BY miv9*-0.066;

SCXIV BY S1XI7*0.617;
SCXIV BY S2XI4*0.447;
SCXIV BY S3XI5@0.724;
SCXIV BY S4XI9*0.645;
SCXIV BY S5XI1*0.613;
SCXIV BY S6XI3*0.443;
SCXIV BY S7XI6*0.551;

SCXUV BY S1XU2*0.438;
SCXUV BY S2XU1*0.330;
SCXUV BY S3XU3*0.468;
SCXUV BY S4XU6*0.314;
SCXUV BY S5XU5@0.497;
SCXUV BY S6XU4*0.465;
SCXUV BY S7XU7*0.309;

 miv with msc (1);
 miv with muv ;
 muv with msc (2);
 edu_asp with mat_ach;
 itsex with her_scl;
 msc with SCXIV SCXUV;
 miv with SCXIV;
 muv with SCXUV;

     MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 WITH
     MIVN2 MIVN3 MSCN2 MSCN3 MSCN7 ;

    S1XI7 WITH MSC1 MIV7;
    S2XI4 WITH MSCN2 MIV4;
    S3XI5 WITH MSCN3 MIV5;
    S4XI9 WITH MSC4 MIV9;
    S5XI1 WITH MSC5 MIV1;
    S6XI3 WITH MSC6 MIVN3;
    S7XI6 WITH MSCN7 MIV6;

    S1XU2 WITH MSC1 MUV2;
    S2XU1 WITH MSCN2 MUV1;
    S3XU3 WITH MSCN3 MUV3;
    S4XU6 WITH MSC4 MUV6;
    S5XU5 WITH MSC5 MUV5;
    S6XU4 WITH MSC6 MUV4;
    S7XU7 WITH MSCN7 MUV7;

     [MIV1 MIVN2 MIVN3 MIV4 MIV5 MIV6 MIV7 MIV8 MIV9];
     [MSC1 MSCN2 MSCN3 MSC4 MSC5 MSC6 MSCN7];
     [MUV1 MUV2 MUV3 MUV4 MUV5 MUV6 MUV7];
```

```
[S1XI7 S2XI4 S3XI5 S4XI9 S5XI1 S6XI3 S7XI6];
[S1XU2 S2XU1 S3XU3 S4XU6 S5XU5 S6XU4 S7XU7];

[MIV@0]; [MSC@0]; [MUV@0]; [SCXIV](1); [SCXUV](2);
EDU_ASP ON SCXIV SCXUV MIV MSC MUV ITSEX HER_SCL;
MAT_ACH ON SCXIV SCXUV MIV MSC MUV ITSEX HER_SCL;
MIV on ITSEX HER_SCL;
MSC on ITSEX HER_SCL;
MUV on ITSEX HER_SCL;

MODEL INDIRECT:
EDU_ASP IND ITSEX;
EDU_ASP IND HER_SCL;
MAT_ACH IND ITSEX;
MAT_ACH IND HER_SCL;

OUTPUT:    STAND CINTERVAL stdyx;
```