THE EFFECT OF CURRICULUM AND CLASSROOM FACTORS ON TIMSS MATHEMATICS SCORES: A CROSS-NATIONAL STUDY

BUKET EREN JANSSEN

BOĞAZİÇİ UNIVERSITY

2017

THE EFFECT OF CURRICULUM AND CLASSROOM FACTORS ON TIMSS MATHEMATICS SCORES: A CROSS-NATIONAL STUDY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Educational Sciences

by

Buket Eren Janssen

Boğaziçi University

2017

The Effect of Curriculum and Classroom Factors

on TIMMS Mathematics Scores: A Cross-National Study

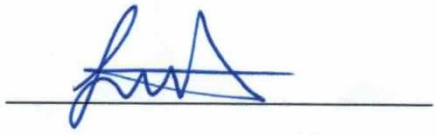
The thesis of Buket Eren Janssen

has been approved by:

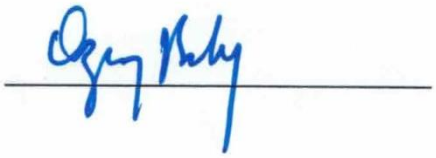
Assist. Prof. Bengü Börkan
(Thesis Advisor)



Assist. Prof. Fatma Aslan-Tutak



Assoc. Prof. Ozan Bakış
(External Member)



August 2017

DECLARATION OF ORIGINALITY

I, Buket Eren Janssen, certify that:

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sorts of ideas and words, including digital resources, which have been produced or published by another person or institution;
- This thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- This is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date.....06.09.2017.....

ABSTRACT

The Effect of Curriculum and Classroom Factors

on TIMSS Mathematics Scores: A Cross-National Study

Mathematics scores differ between countries in international examinations. Factors related to this variation are investigated in various contexts, whereas curriculum related factors have not been considered as part of a comprehensive study until now. Therefore, this study was designed to examine possible factors contributing to variability between countries in Trends in International Mathematics and Science Study (TIMSS) 2015 mathematics test scores. In this correlational research, outcome of interest is TIMSS mathematics scores and predictor variables were extracted from TIMSS student, teacher and curriculum questionnaires. Factors were considered at three levels; country (curriculum), school (classroom context) and student. This study includes 219,788 students in 6,586 schools from 36 countries. Hierarchical linear modeling was selected for analyses due to its usefulness in applying group and individual predictors in their levels simultaneously. Findings show that mathematics scores varied significantly at each level, and total variance was decomposed as 41.1%, 22.8% and 36.0% respectively in student, school and country levels. Results show that countries with more centralized systems and more time assigned on curriculum for mathematics lessons performed lower. More home study support, less working load for teachers, more frequent homework assignments and slightly higher number of students in classes effected to achievement positively. Interactions show countries with more centralized systems, schools with high homework frequency and less number of students performed higher, whereas countries with less centralized systems could achieve higher with more home study support students.

ÖZET

Müfredat ve Sınıf Faktörlerinin TIMSS Matematik Puanlarına Etkileri: Uluslararası Bir Analiz

Uluslararası araştırma sonuçlarına göre matematik puanları ülkeler arası farklılık göstermektedir. Bu farklılığa neden olabilecek etmenler çeşitli bağlamlarda araştırılmıştır, fakat müfredat ile ilgili etmenler şimdiye kadar kapsamlı bir çalışma ile incelenmemiştir. Bu nedenle, bu çalışma Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) 2015 matematik test puanlarındaki ülkeler arası farklılıkta payı olan olası etmenleri incelemek için desenlenmiştir. Bu ilişkisel araştırmada, ilgilenilen çıktı değişkeni TIMSS matematik puanlarıdır ve yordayıcı değişkenler TIMSS öğrenci, öğretmen ve müfredat anketlerinden elde edilmiştir. Etmenler ülke (müfredat), okul (sınıf bağlamı) ve öğrenci olmak üzere üç düzeyde dikkate alınmıştır. Çalışmada 36 ülkeden 6,586 okul ve 219,788 öğrenciye yer verilmiştir. Grup ve birey düzeyindeki yordayıcıları eş zamanlı olarak, düzeylerine göre analiz etme yararından dolayı, çok düzeyli doğrusal model kullanılmıştır. Sonuçlar matematik puanlarının her bir düzeyde farklılaştığını ve toplam varyansın 41.1%, 22.8% ve 36.0% olmak üzere sırasıyla öğrenci, okul ve ülke düzeylerinden kaynaklandığını göstermektedir. Bulgulara göre, daha merkezi bir eğitim sistemi ve müfredatta matematiğe ayrılan sürenin fazla olduğu ülkeler daha düşük performans göstermiştir. Ev öğrenim desteğinin fazla, öğretmenlerin düşük iş yüküne sahip olması, daha sık ev ödevi verilmesi ve öğrenci sayısının biraz yüksek olması başarıyı olumlu etkilemektedir. Etkileşimlere göre, daha merkezi sistemlerde, daha sık ödev verilmesi ve sınıflardaki öğrenci sayısının azalması, daha az merkezi sistemlerde ise öğrencilerin ev öğrenim desteğinin fazla olması başarıyı olumlu etkilemektedir.

ACKNOWLEDGEMENTS

From the very beginning of this thesis writing process, I have been dreaming of this moment to complete the thesis and enjoy the moment of expressing my sincere gratitude to the people who have been there always to support me. Hence, firstly I want to thank the life that it brings me to this very special moment.

The completion of this thesis could not have been imaginable without the courage, sensibility and guidance that my advisor, Assist. Prof. Bengü Börkan has ensured me. So, I would like to express my sincere gratitude to Bengü Börkan. I also want to thank my other thesis committee members for their valuable comments which opened up new horizons for my thesis during the proposal presentation. Thank you very much, Assist. Prof. Fatma Aslan-Tutak and Assoc. Prof. Ozan Bakış.

I also would like to thank to my friends from this master program that I had shared a lot of struggles, hesitations, frustrations etc., but also empathy and attachment together. Thank you all, Neriman Küçüker, Zübeyde Özen, Çağla Pınar, Oğuz Ak and Özkan Alak. All these challenges have made us true friends besides who we are now.

Last but not least, many thanks to my family, especially to my mother Aygün Eren and my brother Barış Eren for their infinite understanding and support during the course period of my master's. I will always be grateful to them. Exclusively, I would like to thank my husband for his permanent support, understanding and exceptional character. Without him celebrations would be deficient. Thank you darling.

Dedicated to

my love Jop and

my little princess Ella Lena

For giving me a constant smile and

strength to follow my dreams ...

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
1.1 The purpose of the study 11
1.2 Research questions 12
1.3 Significance of the study 14
CHAPTER 2: LITERATURE REVIEW 16
2.1 Studies on mathematics achievement as conducting three-level HLM 17
2.2 Studies on students' socio-economic status 20
2.3 Studies related with classroom factors 21
2.4 Studies related with curriculum factors 27
CHAPTER 3: METHODOLOGY 37
3.1 Research design 37
3.2 Participants 37
3.3 Data collection 41
3.4 Instruments 42
3.5 Preparation of data 57
3.6 Data analysis 58
CHAPTER 4: RESULTS 79
4.1 Basic descriptive statistics 79
4.2 Results for unconditional model 83
4.3 Results for only school level predictors included model (Model 1a) 86
4.4 Results for only country level predictors included model (Model 1b) 90
4.5 Results for only student level predictor included model (Model 2) 94
4.6 Results for student and school level predictors included model (Model 3) 98
4.7 Results for final model 103
CHAPTER 5: DISCUSSION 109
5.1 The explained variance in different levels of hierarchical model 109
5.2 The influence of curriculum variables on mathematics achievement 110
5.3 The influence of interaction effects on mathematics achievement 113
5.4 Limitations of the study 116
5.5 Recommendations and suggestions for further research 117
APPENDIX A: CORRELATION ANALYSES 118
APPENDIX B: RESIDUAL FILE ANALYSES 119
APPENDIX C: DESCRIPTIVE STATISTICS FOR RESIDUAL FILES 121
APPENDIX D: EXPLORATORY ANAYSES 124
REFERENCES 125

LIST OF TABLES

Table 1. Numbers of Schools, Classrooms and Students Samples for TIMSS 2015-Eighth Grade 39
Table 2. The Abbreviations of Variables or Indexes 49
Table 3. Recoding for the Index of Teachers' BS Degree 52
Table 4. Recoding for the Teacher Preparedness about Mathematics Topics 53
Table 5. Recoding for the Homework Use in Learning Processes 54
Table 6. Recoding for the Working Load of Teacher 55
Table 7. Hierarchical Identification System Codes Used to Link Schools, Classes, Students and Teachers 57
Table 8. Student-Level Variables (N=219,788) 79
Table 9. School-Level Predictor Variables (N=6,586) 80
Table 10. Country-Level Predictor Variables (N=36) 81
Table 11. HLM3 Analysis Results of Model 1 (Not Any Predictor Included) 83
Table 12. HLM3 Analysis Results of Only Level-2 Predictor Variables Included ... 87
Table 13. HLM3 Analysis Results of Only Level-3 Predictor Variables Included ... 91
Table 14. HLM3 Analysis Results of Model-2 (Only Level-1 Predictor Variables Included) 95
Table 15. Final Estimation of Level-1 and Level-2 Reliabilities 97
Table 16. HLM3 Analysis Results of Model 3-Fixed Effects (Level-1 and Level-2 Variables Included) 99
Table 17. HLM3 Analysis Results of Model 3-Random Effects (Level-1 and Level-2 Variables Included) 101
Table 18. Final Estimation of Level-1 and Level-2 Reliabilities 102

Table 19. HLM3 Analysis Results of Model 4 (Level-1, Level-2 and Level-3 Variables Included) 104
Table 20. HLM3 Analysis Results of Final Model-Random Effects (Level-1, Level-2 and Level-3 Variables Included) 107
Table 21. Final Estimation of Level-1 and Level-2 Reliabilities 108

LIST OF APPENDIX TABLES

Table A1. Correlations Between Level-1 Predictor and Dependent Variables 118
Table A2. Correlations Between Level-2 Predictor Variables 118
Table A3. Correlations Between Level-3 Predictor Variables 118
Table C1. Descriptive Statistics of Level-2 Residual File 121
Table C2. Descriptive Statistics of Level-3 Residual File 121
Table C3. Descriptive Statistics of Level-3 Residual File (Final Model) 123
Table D1. Exploratory Analysis for Estimated Level-2 Coefficients and Standard Errors 124
Table D2. Exploratory Analysis for Estimated Level-3 Coefficients and Standard Errors 124

LIST OF FIGURES

Figure 1. The scheme of intended, implemented and attained curriculum 4

Figure 2. Diagram of the sampling procedures and preparations for the assessment administration implemented by national centers and schools 42

Figure 3. Development scheme of the models 64

LIST OF APPENDIX FIGURES

Figure B1. Normal Q-Q curve of Level-1 residuals for 219,788 students based on the final fitted model 119

Figure B2. Normal Q-Q curve of Level-2 residuals for 6,586 schools based on the interception model 119

Figure B3. Normal Q-Q curve of Level-3 residuals for 36 countries based on the final fitted model..... 120

Figure B4. EB versus OL residuals for the home study support (HSUP) slopes in. 120

ABBREVIATIONS

BS	Bachelor of Science
EB	Empirical Bayes
HLM	Hierarchical Linear Modeling
IEA	International Association for the Evaluation of Educational Achievement
NRC	National Research Coordinator
OLS	Ordinary Least Squares
PIRLS	The Progress in International Reading Literacy Study
PPE	Pre-primary education (Programs including Kindergarten for children age 3 or older)
SMSO	The Survey of Mathematics and Science Opportunities
SES	Socio-Economic Status
TIMSS	Trends in International Mathematics and Science Study
WinW3S	Within-School Sampling Software

CHAPTER 1

INTRODUCTION

In recent years, the world has become a smaller platform because of the technological developments in all areas. Today, traveling to another country, trading with another country and communicating with another country is far easier than the way they had been two decades ago. Because of the opportunities and advantages of the current time, a lot of daily activities have become more comfortable. On the other hand, all these developments make the world a more competitive stage. When these new competitive conditions are considered, as educators, we have to educate our students as more aware of the essential needs of new world as possible.

Education systems have been revised according to the needed characteristics of current-day as also considering the peculiar needs of countries or cultures. As Jacobs (2010) mentioned "Running schools and using curriculum on a constant replay button no longer works" (p.8). Despite the many reform initiatives and expenditures to improve educational outcomes, many educational systems have not achieved the desired performance for years (Schleicher, 2007). One of the aims of an education system might be stated as "ensuring that all students fulfill their potential within the school system" (OECD, 2004, p.8). In order to attain such a goal, it might be beneficial to consider defined goals of an education system, relations between relevant factors of teaching and learning processes and system outcomes together in a global perspective (OECD, 2004). International research results provide policy-makers of participant countries to understand the characteristics of high rank systems and take a lesson from them to adapt constructive and suitable features to their own (Schleicher, 2007).

Curriculum has particular roles in an education system. Curriculum is described as a precise, concrete topics that is always related with decision making processes within organizations either they are schools, nonprofit agencies or governmental institutions (Null, 2011). Curriculum is about which subject should be taught in the classroom environments, but it is not limited only with the subject matter. Additionally, it covers "to whom, how, with what and under what circumstances" the subject matter would be taught. Curriculum associates idea, intention and action (Null, 2011). It is possible to see curriculum as a multifaceted approach in which each facet related with a context or different levels of educational practice (Robitaille, Schmidt, Raizen, McKnight, Britton & Nicol, 1993; as cited in Cogan & Schmidt, 1999).

Curriculum appears in the middle of each contentious argument within schooling and education. As an effort to describe international evaluation standards for curriculum, the *Joint Committee on Standards for Educational Evaluation* (JCSEE) considered five main title of program standards as; utility, feasibility, propriety, accuracy and evaluation accountability (Yardbrough, Shulha, Hopson, & Caruthers, 2011). The described standards are as follows.

Utility standards: This standard is related whether the program is valuable in meeting the needs of stakeholders with the processes and products. It has eighth sub-criteria, which includes the credibility of evaluator credibility, stakeholders, purposes, values, etc.

Feasibility standards: This standard is about efficient and effective evaluation of the program. It has four sub-criteria about the management of a project, procedures, viability and resources.

Propriety standards: This standard is interested in finding the proper and fair ways of program evaluation. It has seven sub-criteria about the quality of orientation, agreements, human rights, and being clear, fair, and transparent etc.

Accuracy standards: This standard is about increasing the precision of the program evaluation, specifically which promote the persuasion about quality. It has eighth sub-criteria generally about reliable information, evaluation, reporting and having accurate program.

Evaluation accountability standards: This standard is focusing on the accountability of program evaluation in a meta-evaluation level. It has three sub-criteria, namely, evaluation documentation, internal meta-evaluation and external meta-evaluation.

These standards can be considered as the accountability standards for any process of education. Generally, educational policies have been decided in the way countries have made considerable amount of expenditure not only as money but also as time in order to create efficient education systems. As an indicator of an efficient education system, mostly countries focus on solely achievement levels and by this way an important point has been missed out; “the process”. Missing “the process” in an education system is alike the saying “No matter how big storms a captain had in a journey, the important point is considered as whether the ship has been arrived at port or not”. As it can be understood from the definition of curriculum in an education system, curricula define the processes. By this perspective, having an effective educational system depends on the transition between theoretic visions of the aims of educational systems and the practical applications in classroom environments (Null, 2011). These transitions can be better explained by intended

curriculum, implemented curriculum and attained curriculum, shown at Figure 1 (Mullis, 2013, p.5).

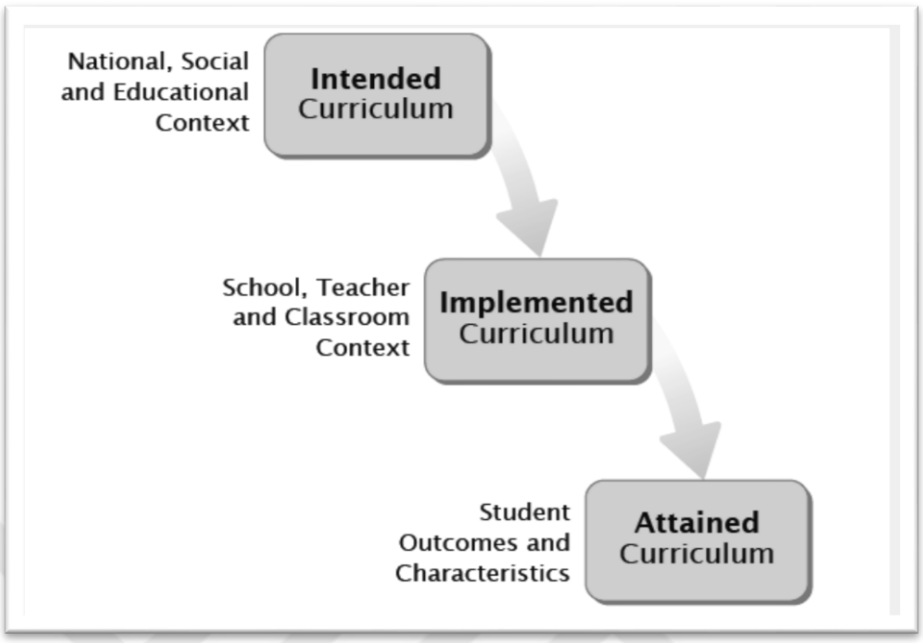


Figure 1. The scheme of intended, implemented and attained curriculum (Mullis, 2013, p.5)

As stated by *National Research Council* intended curriculum refers the expected aims and activities which are planned before the classroom applications. On the other hand, implemented curriculum indicates the curriculum which is experienced by students in the classroom environments. Lastly, attained curriculum is the outcome which resulted from the intended and implemented curricula (National Research Council, 1996). Throughout this thesis, the main focus will be the network between the intended, implemented and attained curricula, because prospective interruptions in the midst of these phases might be the cause of possible problems in an education system. In order to conclude a whole picture about this network, examining it in a cross-national study within the scope of comparative education can be valuable for the future decisions of educational systems.

Comparative education is defined by Mugo and Wolhuter (2013) as “the analytical survey of systems of education across national borders with a view of establishing similarities and differences” (p.4). The comparative education field is also defined as the “intersection of the social sciences, education and cross-national study” (p.2). A lot of cross-national researches which cover range of topics in education have been done by international foundations in order to follow and improve education. The common aims of comparative education and cross-national researches can be stated as follows:

1. Policy makers in individual countries examine education systems in other countries in order to perceive different ways to achieve their political, social and economic objectives.
2. International agencies compare patterns in different countries in order to improve the advice that they give to national governments and others. (Bray, 2005, p. 239)

There are various international organizations and platforms which organize cross-national researches in order to explore the necessary requirements for the new generations to become au fait, competent, innovative human beings in the new era. The *International Association for the Evaluation of Educational Achievement* (IEA) is one of the major independent pioneers in these worldwide agencies. Since IEA was founded in 1958, it has administered large-scale comparative researches about educational achievement and other features of education. It aims to gain comprehensive awareness of the effects of educational policies and practices not only within the systems but also and between different educational systems. As conducting more than 30 cross-national researches which have been conducted, IEA generally aims to (IEA, 2011):

1. Present international measures to determine corresponding strong and weak sides of different education systems.
2. Present data for policymakers in order to understand essential school and extra scholastic facets that effect learning and teaching conditions.
3. Present high-quality data which might be the base of evaluation and development of school systems in different educational systems.
4. Develop or progress the scope of education systems.
5. Contribute to the prosperity of a worldwide researchers' community in educational appraisal.

All these mentioned aims described by IEA are generally focusing on improving the education systems and so the curriculum development ventures of countries, because "curriculum is the heart of education" (Null, 2011, p.1). TIMSS (the *Trends in International Mathematics and Science Study*) and PIRLS (the *Progress in International Reading Literacy Study*) together encompass the core cycle of studies for IEA. TIMSS & PIRLS International Study Center (2013) states purposes of such researches as follows:

> Monitoring system-level achievement trends in a global context, establishing achievement goals and standards for educational improvement, *stimulating curriculum reform*, improving teaching and learning through research and analysis of the data, conducting related studies and training researchers and teachers in assessment and evaluation. (p.1)

All the objectives mentioned above have an effect on curriculum from different perspectives, whereas the objective "stimulating curriculum reform" statement is giving the general idea of the impact of international research results on curricula of each country.

Especially, for last two decades, international tests are in progress to provide data to the cross-national researches on education. These researches provide information not only about achievement levels but also about contextual backgrounds like; house environments, classroom environments, administrative processes, educational policies, etc. The results of these examinations provide information for inquiring into the present circumstances and reviewing the education components such as basic politics in education, curricula, teaching methods, teachers' efficacies and lesson materials.

International studies are done because of a main question in minds as "Why students in a particular country perform better than the students in another country?" The Study of Mathematics and Science Opportunities (SMSO) is a small investigative study which had also been conducted with such a main question (National Research Council, 1996). In order to compare education systems of countries achievement tests are used at that study, because as mentioned at Anastasi (1982) "Achievement tests are designed to measure the effects of a specific program of instruction or training" (p.398). SMSO included six countries which had been foreseen as a representative sample of future TIMSS administrations. This early research was mainly done in order to design the bigger research better as exploring the learning environments in detail. The observations obtained from SMSO might be summarized as follows (National Research Council, 1996):

1. There is not a globally taught set of topics in mathematics or science classes at each grade level.
2. Different countries' curriculum materials diversify regarding the order of subjects, emphasis on different topics and goals for students' achievements.

3. The authority which decides about what and how should be taught, in other words the curricula, varies in different countries.
4. Teaching methods come out very differently across countries.
5. Teaching practices are mostly independent from the culture contrary to the expectation of the research.

Briefly, countries have differences at various stages. They even differ remarkably in how mastery is defined and how the curricula specify that mastery should be achieved (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). Yet they have been compared solely in terms of achievement levels of students. In the line with this thinking, how much accurate would it be to compare countries only based on the achievement levels without considering what differentiate these countries? About this, Isaac Kandel (1881-1965) stated that comparative education is not only about educational structure, organization, curriculum et cetera, but also about finding the causes behind problems of education for different countries and providing solutions for these problems (as cited in Mugo & Wolhuter, 2013).

Therefore, it can be beneficial to evaluate international achievement levels as considering what exactly aimed on curriculum in different countries (intended curriculum), what have been done in classroom environments (implemented curriculum) and how these curricula and actions in classes have an effect upon achievement (attained curriculum). In this research, classroom factors are taken into consideration together with curriculum factors, because of the necessary cooperation between the intended and implemented curricula and their effects on achievement is considered in order to find out the degree of attained curriculum.

TIMSS is an international study that provides information at all three aspects of curriculum. TIMMS had been firstly conducted between 1991 and 1995. Since

then, it gives a trend to participant countries each four years at fourth and eighth grade levels not only for achievement levels of countries in mathematics and science but also factors effect on this achievement, like countries' curricula, classroom environments and teachers' characteristics. Because TIMSS is a globally cooperative program, it has an excessive diversity among the participant countries by means of economic development, geographical location, and population size. In this aspect, one of the main aims of TIMSS is being able to explore factors that influence students' success in different countries (National Research Council, 1996).

TIMSS gives the chance to the participant countries comparing their own achievement levels with each other and learn from the experiences of others by the help of the extra information provided by the wide inclusive database. Participant countries of TIMSS share the common view of comparing their educational organizations, classroom applications, curricula and instructional policies with the rest of the world (TIMSS and PIRLS, 2013). TIMSS used curriculum variables and some other contextual variables in order to interpret in what way educational opportunities are given to students and how students use these opportunities (Mullis, 2013). TIMSS 2015 is the sixth research in IEA's 20 years series of international assessment which aims to improve educational processes in mathematics and science with a worldwide perspective.

TIMSS includes background questionnaires as student, teacher, school and curriculum questionnaires, besides mathematics and science achievement tests (Martin, Mullis & Foy, 2013). Besides the curriculum and classroom variables which effect students' mathematics achievements, students' socio-economic status (SES) is also a very important factor to understand the family effects on achievement. Based on various research results, especially if SES is considered as an individual factor or

an aggregated factor, the effects of SES on achievement varies (White, 1982; Sirin, 2005). For example, White (2005) mentions on his very first meta-analytic research that SES is very weakly correlated with achievement on individual level (r=.22), whereas it has higher correlation on aggregated levels (r=.73). Hence, while investigating the factors at different levels, examining an indicator of students' socio-economic status at student level might also be an important aspect of the current study. Therefore, data gathered by curriculum questionnaire, teacher questionnaire, student questionnaire, and mathematics test is used in this study in order to comprise the three aspects of curriculum.

For most studies, it is not possible to determine how much the intended, implemented and attained curricula correspond to each other (Mullis, Martin, Foy & Arora, 2012). There are researches which considered variables at class level for instance teacher characteristics, such as preparedness, certification, working load (Mandeville and Liu, 1997; Darling-Hammond, 2000; Boscardin et al., 2005) or other classroom characteristics such as class size, homework frequency (U.S. Department of Education [USDE], 1986; Lamb & Fullarton, 2002). In most researches, these characteristics are considered without taking account the hierarchical structure of the data or considered together only with student variables as two level. As SMSO mentioned educational process is affected at three levels as system, classroom and student (Cogan & Schmidt, 1999). Although there are researches which investigate educational processes in three levels (Chiu; 2010; Mohammadpour & Mohamed Najib, 2014; Çelik, 2016), there has not been found a research which has considered curriculum variables at country level in order to understand the effect of curriculum variables. Therefore, in this study, it is aimed to examine how the variation in eighth grade students' mathematics achievement scores

in TIMSS 2015 participating countries is distributed at student, school, and country levels.

1.1 The purpose of the study

Previous comparative researches show that there is variability between participating countries achievement scores. Differences between countries were explained mostly by student, classroom, school and country variables in these researches. As well as these variables, the curriculum characteristics of the countries are expected to contribute students' achievement scores. Comparative researches are an opportunity for us to evaluate the effect of different policies on curriculum and classroom contexts on TIMSS mathematics test scores. It is important to understand which factors related to curriculum, classroom and students have impact on students' achievement scores, and how these factors interact with each other to effect students' scores as well. Therefore, in this comparative study, the purpose is to reveal the curriculum, classroom[1] and student related factors behind the variation of mathematics achievement scores by the TIMSS eighth grade participating students. It is also aimed to investigate the cross-level interactions and their impacts on mathematics achievement scores at student, school and country levels. In TIMSS 2015 administration, the data shows a strictly hierarchical structure as with schools nested within countries, classes nested within schools and students nested within classes (Johansone, 2016). Because of the structure of the data and the advantages hierarchical linear modeling provides, hierarchical linear modeling is used for analyses.

[1] Due to some data related issues, variables about classroom context were taken into the analysis as aggregated at school level, therefore throughout the text, it is referred as 'classroom context' to mention the characteristics of the classroom environment but also referred as 'school level' to mention the statistical level of the variables.

The outcome variable of this study is the mathematics achievement scores of eighth grade students. In the line of the research purpose, the predictor variables at country level are the indexes of; components prescribed by mathematics curriculum, methods used to evaluate the implementation of mathematics curriculum and variables of; instructional time devoted to mathematics curriculum, number of TIMSS topics intended to be taught by the end of eighth grade, mathematics and numeracy skills at the pre-primary education and classroom tracking/grouping by achievement. Predictor variables at school level are the indexes of; percentage of students taught the topic 'Numbers', percentage of students taught the topic 'Algebra', percentage of students taught the topic 'Geometry', percentage of students taught the topic 'Data and Chance', teacher preparedness about mathematics topics and homework use in learning processes and variables of; teachers' BS degree, working load of teacher, number of student in the class, weekly mathematics class time and homework frequency. Lastly, predictor variable at student level is included as home study support. At following sections, these mentioned variables will be given with their operational definitions and will also be given in a table for clarity.

1.2 Research questions

The main research question that guides this study is: "As considering the participant countries attended to TIMSS 2015, what are the country level curriculum variables that effect 8^{th} grade students' mathematics achievements and how student and school level predictor variables are related with that affect?"

In the direction of the main research question, following sub-questions are examined:

1. Are there significant differences between students' average mathematics achievement scores on country level and how is the variation in mathematics achievement scores distributed at student, school and country levels? (Model 1: Unconditional Model)
2. If there is a significant difference between countries' average mathematics achievement, to what extent does each of the school level curriculum variables explain that difference when student and country level predictor variables are considered as unconditional? (Model 1a)
3. If there is a significant difference between countries' average mathematics achievement, to what extent does each of the country level curriculum variables explain that difference when student and school level predictor variables are considered as unconditional? (Model 1b)
4. What is the effect of the student level predictor variable, home study support on students' mathematics achievement? (Model 2)
5. How does the effect of student level predictor variable on students' mathematics achievement depend on the school level predictor variables? (Model 3)
6. What is the overall effect of the predictor variables on mathematics achievement scores, when student, school and country level predictor variables considered together? (Model 4: Final Model)

1.3 Significance of the study

Nowadays, countries have had the chance to evaluate their students' achievement levels by the help of international surveys. However, learning one country's students' higher scores compare to other country's students may give us the fact but not the reasons behind that fact. In TIMSS, in order to understand the background reasons of achievement levels, context questionnaires are applied to the related sample groups. When answers to these context questionnaires are compared across countries and analyze in relation to achievement levels, this knowledge might provide intuition for a bigger picture of effective educational policies. Because of the fact that, the current study provides a three level model including student, classroom and curriculum context variables, it is believed that the results might give some tips about the factors which related with mathematics achievement.

One of the impressive data at TIMSS is the data gathered by the curriculum questionnaire. Evaluating those data might provide a perspective to the low achieving countries on mathematics tests about the curriculum of counties in order to improve their mathematics scores. To the extent of literature review, there has not been a research done as using curriculum data of an international research. Therefore, it is thought this study can make a difference due to the fact that its being the first research which uses information about curriculum from many countries of different characteristics.

There are always differences between the intended, implemented and attained curricula, which mean that what is aimed and what comes true, might be different frequently. Because of that fact, examining what is happening in the classroom environments by the help of classroom contexts can give opportunity for further comments. Therefore, examining mathematics curriculum data and classroom

contextual data comparison with mathematics achievement levels of countries can give an important inspiration to a country's policy makers. Furthermore, investigating the relations of factors and the interaction effects of factors on mathematics achievement might also give ideas about some conditional situations. For instance, taking students' socio-economic status into these interaction effects might also show the conditions linked to the family background of students. Especially, it is important to consider how a child's socio-economic status is related to a country's curriculum, in order to understand if the interaction of a child's socio-economic status and a curriculum factor effect to student's achievement level.

Finally, exploring and evaluating the results of large scale tests such as TIMSS, might offer information to evaluate Turkey's own curriculum along with other countries (Akyüz, 2006). As stated at Mullis et al. (2012) the information gathered by such international surveys, "may be used not only to help understand the relationships among policies, practices, and outcomes within countries, but also to provide potentially powerful insights into international best practice" (p.1). At the end, reviewing other TIMSS countries' curricula in mathematics education aspect might give an inspiration to interpret Turkey's education curricula. With this thesis, it is hoped to provide a growth model perspective to schools systems which is positive and deep, instead of a change model perspective, which can mostly be perceived as a trendy approach.

CHAPTER 2

LITERATURE REVIEW

In this part, the researches findings inquire student, classroom and curriculum factors which are summarized in the scope of this study as related with mathematics achievement.

Firstly, it would be congruent to give some information about the *Survey of Mathematics and Science Opportunities* (SMSO) which laid the foundation of TIMSS. SMSO aimed to develop curriculum frameworks for both mathematics and science areas, a methodology for an in-depth analysis of curriculum, improved measures of instructional practices and students' opportunity to learn specific curricular topics (Cogan & Schmidt, 1999, p.69). In that project, observations were undertaken in order to uncover instructional practices that held promising potential to explain international variation in student achievement and to clarify what occurs in classrooms as intended curriculum goals are transformed into potential experiences for students. According to the model explained at SMSO, various factors affect the educational processes at three different levels, which are system level, classroom level and student level. (Robitaille, Schmidt, Raizen, McKnight, Brittan and Nicol, 1993; as cited in Cogan & Schmidt, 1999).

Therefore, in a similar vein with mentioned study SMSO these three levels are also the subject of the current study and literature review is presented in three titles as curriculum related factors, classroom related factors and students' socio-economic status. It should be kept in mind that in this current study, the effects on mathematics achievement scores of student and school level classroom

characteristics are examined, but these are basically considered as control variables and included to see interaction effects with country level curriculum variables.

2.1 Studies on mathematics achievement as conducting three-level HLM

Strengthening students' achievement levels in mathematics is routinely a prominent issue of worldwide educational reform efforts. International studies focus on mathematics achievements of countries and the factors which effect mathematics achievements. It is critical particularly considering the essential need of mathematics knowledge for further innovative engineering progresses, which is one of mandatory requirements of developing world. When considering the factors which affect mathematics achievement, it would be a mistake not considering the structure of educational systems. Therefore, researches which considered the hierarchical structure of educational systems, such as students nested in classrooms, classrooms nested in schools and schools nested in countries, would be necessary to mention.

Çelik (2016) investigated the country level factors in the scope of TIMSS 2011 mathematics achievement scores. The findings show that the total variance in mathematics scores were shared as 42.8%, 24.9% and 32.3% on student, school and country levels, respectively. At the end of conducting three-level HLM when including only country level predictor variables, it is concluded that "expected years of schooling", "educational attainment at least completed tertiary" and "grades of compulsory education" explained around 79% of the variance in TIMSS mathematics scores at country level.

Mohammadpour and Mohamed Najib (2014) examined three-level HLM analysis in order to observe how the variation in TIMSS mathematics scores dispersed at country, school and student levels, and how the achievement and

different factors at each level are linked to each other. Furthermore, they also examine the origin of the variation in mathematics achievement within every participant country. According to the results, 40.39% of the variance accounted for within-school differences, whereas 20.61% for between schools within country and 38.99% for between countries. Students' socio-economic status appears as the second strongest predictor at student level after mathematics self-concept, where a one scale point increase of socio-economic status increases the mathematics achievement scores 17.57 and only socio-economic status explained 2.84% of the variance. Furthermore, students from urban schools performed higher than rural schools. The schools which have positively described school-climates and larger-sized classrooms attained a higher average on mathematics achievements than their counterparts. At country level, country socio-economic status is found as the most critical predictor of the average mathematics achievement level.

Chiu (2010) studied how country, family and school features were connected to mathematics achievement scores. According to the student, school and country level analysis Chiu (2010) stated that equal dissemination of resources promotes general student learning. Furthermore, students in families and schools with more resources (high SES, physical possessions like books, computer etc., and teachers with good education) had higher grades than their peers who had limited resources. On the other hand, intangible processes like affective factors were more important for students in richer countries, who already had physical resources publicly.

Lamb and Fullarton (2002) investigated the reciprocal relations among various variables in US and Australia at three different levels as student, classroom and school. The main aim of that study was to reveal the teacher factors which shape the differences in students' mathematics achievements rather than classroom effects.

Hence, the policy makers could focus on teacher trainings more if it was found necessary. According to the findings, school level socio-economic status and tracking the students according to their ability level heavily affect students' mathematics achievements. Mathematics achievement was pointed out highest for the students from middle socio-economic status families and grouped at highest track. One of the remarkable finding of that study was tracking worked well for highest ability group students whereas bottom group students performed worse than their peers at schools which did not have tracking according to ability.

Ryoo (2001) examined the multi-level effects on students' mathematics sores as using TIMSS 1995 data. The sample of the study was formed by 35 country included 3,999 schools and 128,231 students. According to the results of three-level HLM analysis, 69.29% of the total variance is sourced by within-school differences, 11.51% is sourced by between-school differences and 19.19% is sourced by between-country differences. Additionally, the analyses were conducted to reveal the predictors of these differences at each level. According to the findings 9.34% of student level variance is explained by the variables of students' socio-economic status, language, extra mathematics classes, gender and not being willing to study. On the other hand, 40.02% of the school level variance is explained by the variables of teacher-student ratio, teacher experience, average socio-economic status, average working hours and tracking according to mathematics ability. Lastly, 33.32% of the country level variance is explained by the variables of national examination policy and gross national product (as cited in Çelik, 2016, p.14)

2.2 Studies on students' socio-economic status

On mathematics achievement, most of the variance comes from the student level predictor variables (Ryoo, 2001; Lamb & Fullarton, 2002; Chepete, 2008; Mohammadpour & Mohamed Najib, 2014; Çelik, 2016), whereas within the scope of the current study, it is focused more on country level curriculum variables. Hence, student level predictor variables are limited with one variable which is considered as an indicator of students' socio-economic status.

In TIMSS administrations, there are different items to question the socio-economic status of students (e.g. Do you have any of these things in your home? Your own room; about how many books are there in your home? What is the highest level of education completed by your mother/father?) (Foy, 2017, *Supplement 1*). There are also composite index variables develop by IEA. Two of the index variables are about students' socio-economic status. These variables are "number of home study support" and "parents' highest education level" (Foy, 2017, *Supplement 3*). In this current study, only the number of home study support composite variable is included into the study, due to the possibilities available. The amount of missing data for "parents' highest education level" variable is more than the decided criteria of 5% (Tabachnick & Fidell, 2001). Therefore, students' socio-economic status is tried to be represented by number of home study support.

In literature, numerous indicators of socio-economic status (e.g. number of books, family wealth, parents' education etc.) had been used to predict students' achievements. Mostly, the directions of the relation were similar whereas when the number of indicators became higher the strength of the relation would also be higher (White, 1982; Yang, 2003; Chepete, 2008). For instance, Yang (2003) and O'Dwyer (2005) included different socio-economic status indicators in their studies yet they

found very similar findings. Hence, it is understandable that students who had more books at their home got mostly higher scores in mathematics than their peers who had less (Chiu, 2010) or home-possessions and frequency of speaking English at home have positive and statistically significant relations with mathematics achievement levels (Chepete, 2008).

2.3 Studies related with classroom factors

The researches at this part are about the classroom related variables which also includes variables about mathematics teachers. These variables are percentage of students who are taught mathematics topics of numbers, algebra, geometry, and data and chance topics in the classroom applications, teachers' preparedness to teach mathematics, homework use in the learning processes, teachers' BS degree, working load of teachers, number of students in the classes, weekly mathematics class time and homework frequency. Hence, at this part of the literature review, the research summaries are given around the related researches about these variables.

2.3.1 Studies related with the percentage of students taught mathematics topics

Arora (2010) mentioned at his/her study that one of the most important factor which predicted students' achievement levels was the percentage of students who were taught the topic according to their teachers' answers. The composed variables about percentage of students taught on mentioned topics were derived from items which were about whether the content coverage provided for students or not in classroom environments. Hence, the content coverage in classroom environment is also related with this part of literature.

Boscardin, Aguirre-Munoz, Stoker, Kim, Kim, and Lee (2005) examined whether and how different opportunity to learn variables effected on student outcomes and if there was a consistency across different content areas. One of the statistically significant finding was “teachers who spent more time covering assessment content areas… have higher average student test scores” (p.323).

2.3.2 Studies on teachers’ characteristics

The study which was conducted by Boscardin et al. (2005) aimed to figure out how variables of opportunity affected students learning outcomes. According to the results, teacher expertise and knowledge on the content areas have positive relation with achievement levels of students. Additionally, Arora (2010) stated that one of the most important factor predicted students’ achievement levels is teachers’ reports of their level of preparedness to teach the topic. Carroll (1989) also mentioned that “teachers must be not only intelligent and competent as classroom managers but also adequately knowledgeable about the subject matter they teach”. (p.30)

Hill, Sleep, Lewis and Ball (2007) focused on the assessment of mathematics teachers’ subject knowledge and made a discussion about the importance of the subject knowledge and pedagogical action to transfer that content knowledge to students. At the end, they concluded that giving motivation of learning to students, organizing the learning environment and organizing the information given to students about content were important as well as content knowledge.

Mandeville and Liu (1997) examined student performance according to the interaction effect of teacher preparation on mathematics and thinking level of mathematics problems by their study. They used 118 mathematics items which included from lower to higher thinking ability as 43 knowledge and recognition

items, 52 understanding items and 23 thinking skill items. According to the findings, student performance in open-ended items which were considered as high level mathematics problems differentiates the students' achievement levels depended statistically significant on teachers' qualifications like level of certification and teacher preparation on mathematics topics.

On the other hand, as also found by Monk (1994) content preparation has a positive relation with student achievement level till to a threshold value. Because of a curvilinear relation, student achievement started to decrease after that threshold point (as cited in Darling-Hammond, 2000). This curvilinear can be considered as the unclear research results about the relationship between teacher preparedness of subject matter and students' achievement levels.

Chepete (2008) examined mathematics achievement in schools in Botswana as conducting two-level hierarchical linear modeling at student and school levels. One of the most noticeable findings of that study was the higher correlation between students' achievement and having an education major instead of a certification in mathematics. This finding supports that having a major in mathematics education improving the mathematics achievement levels of students more than having a major on only mathematics.

Darling-Hammond (2000) examined the relations between teacher qualifications and other classroom and school inputs with student achievement levels with aggregated data at state level. She had used a data set from 1993-94 administration of Schools and Staffing Surveys together with the National Assessment of Educational Progress belonged to 50 states. According to the findings, teacher qualifications like certification degree in the field to be taught (have a certificate on education area) had very significant and positive correlation with

student outcomes, namely achievement. Teacher qualification appeared as the most persistent and highly significant predictor of student achievement. Students from teachers who were fully certified and had a major in the field they teach were the most successful students in reading and mathematics (Darling-Hammond, 2000). On the other hand, study conducted by Lamb and Fullarton (2001) stated that teachers' educational qualifications do not have a significant effects on student achievement. Similarly, Atar (2014) also stated that from teacher qualifications, the faculty to be graduated did not have a significant effect on science achievement scores.

On the other hand, as Bandura (1993) mentioned shared efficiency of teachers within a school is one of the most dominant variables that differs among schools and orderly connected with student achievement (as cited in Goddard, Hoy, & Hoy, 2000). Furthermore, Firestone and Pennell (1993) emphasized that high teacher commitment also increases student achievement. Teachers with professional commitment are at a lower risk of psychological deprivation and so these teachers can be more sympathetic toward students, have a higher tolerance for frustration in the classroom, and feel less anxious (Firestone & Pennell, 1993). In order to provide efficiency and commitment of teachers, it is important to provide more favorable working conditions and less working load out of classroom as possible. In other words, teachers' working conditions matter a lot not only to them but also to their students (Johnson, Kraft, & Papay, 2011). Therefore, it is suggested policy-makers to consider teachers' working conditions and the importance of teachers in planned educational reforms (Bascia & Rottmann, 2011).

2.3.3 Studies related with homework frequency and using reason of homework in learning processes

According to the John Hattie (2009)'s study results, the most important classroom variables are proper feedback and active people in learning process. Assigning homework for following classes and checking the answers in order to understand what was understood well during previous lesson and what was understood less would provide effective feedback not only for teachers but also for students.

Lamb and Fullarton (2002) stated that time spent on mathematics homework had strong predictive effect on mathematics achievement in both countries considered, especially in Australia (considered with US). They concluded that "classes where teachers set more homework were associated with higher levels of achievement" (p.167).

In 1986 report by U.S. Department of Education, findings showed about the quantity of homework that "student achievement rises significantly when teachers regularly assign homework and students conscientiously do it" (p.41). This finding was verified for reading, science and mathematics. It is reported that homework increased achievement generally because the time spent on task became higher. But of course the amount of given homework is not the only criterion to be provided for higher achievement levels of students, but also quality to be considered (U.S. Department of Education, 1986).

Quality of homework is mentioned on the same report as, "Well-designed homework assignments relate directly to classwork and extend students' learning beyond the classroom. Homework is most useful when teachers carefully prepare the assignment, thoroughly explain it, and give prompt comments and criticism when work is completed" (USDE, 1986, p.42). Homework needs to be considered as

important by teachers and accepted as a part of the instruction, and eventually have to be discussed after completion in order to support lesson. Shortly, homework assignments have to be used effectively by teachers to direct their students to be independent learners.

2.3.4 Studies on class size

Lamb and Fullarton (2002) considered class size at school level as a variable in their study in order to consider the effect on achievement level. According to findings, average class size of schools did not have a significant effect on achievement and explained very small variance in total variance. Despite the effects is not significant, class size has negative (for US -0.9, for Australia -0.4) relation with achievement at school level for both countries.

Summers and Wolfe (1975) mentioned about school size which also stated to be in a more or less crowded groups in learning environments. They suggested smaller school sizes made a positive change on learning at the elementary and senior level schools.

Eberts, Kehoe and Stone (1984) investigated whether school size is effecting to student mathematics achievement because of its impact on quality and quantity of instructional time, besides the exploration of the relationship between school size and some school factors which were about administrative and instructional processes. According to the results, large population schools were found significantly less effective for student achievement. That result assumed to be linked with differences in resources between small and large population schools (Eberts, Kehoe & Stone, 1984). On the other hand, Darling-Hammond (2000) and Atar (2014) stated that class size did not show a significant relationship with student achievement.

2.4 Studies related with curriculum factors

The researches at this part are about curriculum related variables considered this study. These variables are components of curriculum, instructional time devoted to mathematics, coverage of mathematics topics, evaluation ways of the curriculum, mathematics and numeracy skills at pre-primary education and classroom tracking by achievement.

2.4.1 Studies about components prescribed by curriculum

The concept of components prescribed by curriculum mentioned whether the curriculum of a country specifies goals and objectives, instructional processes and methods, materials and methods of assessment. Nunan (1988) stated the key elements of curriculum as planning, content selection (deciding the objectives), methodology (which includes selection of education activities and materials), and evaluation during the ongoing activities. In order to mention the importance of defining the components of a curriculum, Su (2012) stated that the way a curriculum was designed also included the destination whether there would be a success and what curricular implementation would be obtained.

Hattie (2009) had concluded a comprehensive study as doing meta-analysis of more than 800 meta-analysis studies relating to achievement. In consequence of analysis, 138 impacts are detected about student, teacher, school, curricula or home variables. Some of the curricula variables which have been appeared with high impact sizes might be listed as; explicit and appropriately challenging goals, evaluation of whether the goal is attained and conscious practices to attain to the goals. Therefore, Hattie (2009) presents the importance of prescribed components of

curriculum, namely goals, instructional processes or method and materials, and assessment methods.

At this point, whereas defining the goals, processes, materials and evaluation methods has beneficial sides such as giving the same opportunity to all over the country or state, this is an issue of how centralized curriculum components should be defined. Centralized systems are being supported especially by some countries which are proponent of "unity, integration and strong central control" (Rado, 2010, p.12), because of developmental reasons mostly. As Rado (2010) mentioned, mainly "developing countries" follow "centralized education governance systems" (p.12). On the other hand, systems which follow decentralization in educational processes are sharing the power and responsibilities with local parties such as school administrations. Rado (2010) criticized centralized educational systems as mentioning two major limits. One of these limits is the inadequate involvement of the essential shareholders in decision making processes (Rado, 2010). This inadequate involvement causes the administrative power not noticing the experiences, view and inoperative sides of the applied curricula, which mean the gap between the intended and practiced would be large. The second deficiency is the central power is limited to direct all implications in the schools or classrooms, whereas for instance school administrative can be more aware of all process, what works or what need adjustments. All educational systems are aiming to provide the intended features of effective teaching to improve students' learning. Rado (2010) emphasizes "school-based quality management" (p.168), especially for closing the gap between the intended and implemented curricula.

2.4.2 Studies about instructional time

Berliner (1990) stated instructional time as a multifaceted concept which was very useful and exploratory because of its power to understand, predict and control the process of learning. Gettinger (1984) concluded the importance of time in education as "…time is an important variable in learning that students differ in the amount of time needed to learn…" (p.25). On the other hand, some researchers stated instructional time as a trivial variable (McNamara, 1981). For instance, according to the Heyneman and Loxley (1983)'s study as having same instructional time, students from low per capita countries had learned less than students from high per capita countries, which meant instructional time and its effects change according to the income level of the country.

Instructional time is considered as a superior concept which comprised many terms, such as allocated time, time-on-task, engaged time, transition time, academic learning time, waiting time etc. (Berliner, 1990). All mentioned branches of instructional time have different importance during school learning. Within the scope of the current study, mentioned concept is 'allocated time' which is the provided or scheduled time for students by the national ministry, school or teacher.

Allocated time is also mentioned in literature as "opportunity to learn" and defined as "the amount of time allowed for learning" (Carroll, 1989, p.26). When instructional time is mentioned at literature, mostly aptitude is also mentioned with it based on Carroll's model (Carroll, 1989). Aptitude was defined as "the amount of time a student needs to learn a given task, unit of instruction, or curriculum…" (p.26). Furthermore, it is strongly mentioned that students differ in terms of the amount of time they need to learn a specific task and this situation also changes from

topic to topic (Gettinger, 1984). Hence, in order to get higher achievement, students' aptitudes and given opportunity to learn for students need to be matched.

Allocated time is linked to the determined length of the school year, school day and the number of grades to instruct on decided topic. On the other hand, the indicators which are closely related with instructional productivity, such as absenteeism rates of students or teachers, dropout rates, summer school or after school supports are not included in that concept for the current study. These latter mentioned variables are considered mostly in the scope of time-on-task measures (Berliner, 1990). Allocated time is considered as a variable which can be controlled by policy makers at state or country levels whereas time-on-task cannot be controlled easily by country level mechanism. On the other hand, allocated time is considered as a weaker predictor of achievement than time-on task whereas time-on task is harder to determine (Berliner, 1990). It is also emphasized that allocated time and time spent on particular topics in classroom environments varied a lot (Gettinger, 1984). Yet allocated time decides for students the amount of opportunity to learn and very important to reveal its impact on achievement levels of students.

2.4.3 Studies about curriculum alignment

The intended mathematics curriculum differs between countries, not only about how curricular goals are given in national or school level documents but also about the way curricular goals follow the changing advances (Hooper, Mullis & Martin, 2013). Hence, examining the alignment of curriculum objectives of countries and the assessed objectives in TIMSS administration is of capital importance.

Mitchell (1999)'s study aims to examine the relation of curriculum alignment and student mathematics achievement. Curriculum alignment is explained as the

match between written, taught and tested curricula. The main idea of the study is "what is taught should be tested and what is tested should be what is taught". (p.3) Results of the study shows that curriculum alignment occurs as an important factor to increase achievement of students.

A similar study is done by Squires (2012) and results show a similar conclusion that schools can improve the achievement levels of their students as paying attention to the match between what is aimed and what comes true. In order to do that, curricula can be investigated with the perspectives of its small components, coverage, the timing and evaluation processes.

Cal (2011) conducted a study which aimed to study the level of alignment of the learning outcomes, textbooks and national evaluation process in Belize. The findings of this study mentioned the goals of a curriculum should be specified adequately by curriculum developers, which means goals had to be accurate for effective learning process and higher achievement levels. Furthermore, these goals had to be matched with what would be evaluated.

In addition, Boscardin et al. (2005) stated that content coverage in classroom environment affects students' performance positively, which is a consistent result with the explanations of Porter (2002)'s study.

Arora (2010) had aimed to determine how student performance across large content domains differs in relation to curricular coverage and instructional effectiveness. One of the important results, Arora (2010) mentioned that each country's achievement is relevant to "whether the topic was included in the country's curriculum" (p.1). The findings of this study highlighted one more time that curriculum is a phenomenon which can be controlled by countries.

2.4.4 Studies about evaluation methods of the curriculum implementation

The evaluation of curriculum implications is done mostly by national committees at highly centralized systems, whereas at decentralized systems evaluation processes of implications are mostly assigned to the school administrations or teacher committees. At centralized systems, the schools are expected to be similar with each, in terms of curriculum, textbooks as well as assessment methods (Hooper, Mullis & Martin, 2013). Therefore, the evaluation methods of the curriculum implementation are taken form according to the educational system of countries.

Van de Werfhorst and Mijs (2010) revised comparative education researches as focusing on national-level institutions and possible inequalities in student achievements. According to the findings, they stated that more centralized assessments of implications provided less inequality of opportunity for students compared with more differentiated educational systems.

Bishop and Woesmann (2004) investigated educational production efficiency through the impacts on student performance, such as central examinations, central versus school autonomy, etc. According to the results, positive effects were observed for central examinations and centralization of standard settings and performance control instead of decentralized systems. In other words, in terms of student performance which is affected by the curriculum implication, the system needs to have a check mechanism which should be standardized and centralized.

Bol, Witschge, Van de Werfhorst and Dronkers (2013) studied the effect of central examinations on the collaboration of tracking and inequality. Findings showed that in educational systems with central examinations, parent's social class had smaller effect. This relation was linked to accountability impact for schools

through centralized examinations. In other words, it is suggested that centralized examinations provide more school level accountability for educational systems.

On the other hand, there were also researches which had no clear evidence on the relation between differentiation of achievement tests and average achievement of countries (e.g. Van de Werfhorst & Mijs, 2010). Moreover, according to Chepete (2008)'s findings, schools showed statistically significant differences in even centralized systems.

Ruiz-Primo, Shavelson, Hamilton and Klein (2002) introduced an approach to assess the effect of education reform on student achievement as using the evaluation practice from the closest to student (which were students' products) to distant from student (which were national exam applications). They used close, proximal and distal performance achievement and according to the results, it was found parallel to the expectation that close assessment was more sensitive to the changes in student performance whereas when it becomes further the sensitivity became decreased. The impacts of educational reforms are mostly not observable according to the assessment results. Evaluating students' performance in national level instead student level might be the reason of those discouraging results.

Therefore, national examinations give more accountability and equality opportunity to systems, whereas they are less sensitive. On the other hand, only school level evaluation caused inequalities between schools and provides less accountability of schools, whereas school level evaluations are more sensitive for individual differences. Hence, both centralized and decentralized evaluation methods have advantages and disadvantages. Accordingly, it might be concluded that the more evaluation methods are used to check curriculum implications than not only more accountable but also more sensitive systems can obtained.

2.4.5 Studies about mathematics and numeracy skills at pre-primary education

Research evidences state that pre-primary education is important for future academic performance of children (Taiwo & Tyolo, 2002; Berlinski, Galiani & Gertler, 2009; Tucker-Drop, 2012). For instance, Taiwo and Tyolo (2002) compared Botswana grade one students with pre-school experiences and without pre-school experiences in terms of English language, mathematics and science performances. According to the findings, they stated that grade one students who had pre-school experience performed higher than in all three subject areas. That result was also verified for boys and girls separately.

TIMSS 2011 Encyclopedia (Mullis et al., 2012) mentioned that countries differ from each other seriously in their pre-primary education policies and routines. There are some examples of research findings which supported the positive effect of pre-primary school on academic achievement at primary school are given as Berlinski, Galiani, and Gertler (2009) and Tucker-Drop (2012) (as cited in Hooper, Mullis & Martin, 2013). These researches are summarized below.

Berlinski, Galiani and Gertler (2009) investigated the effect of pre-primary education on third grade at primary school performance of Argentinian students. According to the findings, pre-school education increased not only primary school test scores but also had positive effect on behaviors like attention, class participation etc.

Tucker-Drop (2012) suggested that preschool may diminish the achievement gaps between the socially advantaged and disadvantaged families. The major finding was interpreted as “preschool attendance was prospectively associated with enhanced reading and math skills, particularly for racial and ethnic minorities, children from lower-SES families…” (p.7). In other words, it is concluded that pre-school is

crucially important for academic achievement in further grades and this effect is higher for socially disadvantaged groups.

On the other hand, for very young children's mathematical and numeracy skills have been also considered with specific researches (e.g. Van de Rijt et al., 2003). The research conducted with Belgium, England, Germany, Greece, Slovenia and The Netherlands in order to compare the progress of children with changing social, cultural and educational experiences together with variables such as age, sex and school experience. According to the result, surprisingly there were trivial differences stated between countries despite the severe differences of schooling experience (Van de Rijt et al., 2003). Hence, the findings of this study showed mathematic or numeracy education at pre-primary education did not provide a significant difference.

2.4.6 Studies about classroom tracking/grouping by achievement

Tracking or grouping according to ability or achievement levels of students is very popular especially some countries, such as US. This practice has some negative impacts as well as its positive effects. Lamb and Fullarton (2002) and Van de Werfhorst and Mijs (2010) stated that both in US and Australia tracking has a big effect on mathematics achievement at classroom level. The effects of tracking is summarized as, "Tracking or streaming clearly benefits those students in the higher band classes, but leads to significantly poorer achievement in lower band classes" (Lamb & Fullarton, 2002, p.167).

Schütz, Ursprung and Wössmann (2008) examined the students' educational performance based on family background impact and they stated that the variation was related to some systematics characteristics of the countries' education systems.

The most important systematic features were obtained as pre-school education had to be broad and tracking had to be at late age groups.

Furthermore, Hanushek and Wössmann (2006) examined the student outcomes of tracked and non-tracked systems for primary and secondary school levels. According to the results, tracking especially at early ages increased inequality in education. There was also an unclear tendency obtained that tracking to decrease average performances.

Ho (2009) suggested in her research about tracking and streaming as practices which need to be examined more detailed in order to understand both convergent and divergent characteristics of learners from East Asian who have been outperformed their peers from Western countries. It is especially mentioned to consider tracking effect on TIMSS' eighth grade results (Hooper, Mullis & Martin, 2013).

CHAPTER 3

METHODOLOGY

In this chapter of this thesis; research design, participants, data collection, instruments, preparation of data and data analysis of the thesis are presented.

3.1 Research design

The aim of this study is to determine if Trends in International Mathematics and Science Study (TIMSS) 2015 eighth grade mathematics achievement scores vary between countries and if there is a variation between countries, to determine the curriculum, classroom and student related factors behind that variation. In order to attain this goal, this study is designed as a correlational research, due to the fact that the correlational studies aim either to determine relationships between variables or use these relationships to make predictions for the likely outcomes (Frankel, Wallen & Hyun, 2012). Those studies provide a numerical estimate of how the variables are related, and the higher this numerical estimation is, the higher the relation or prediction power is accepted (Gay, Mills & Airasian, 2006).

3.2 Participants

The target group taken within this research is the eighth grade target group of TIMSS 2015 administration. TIMSS administrations have two target groups in two different grade levels, namely 4^{th} and 8^{th} grades. These target populations are defined according to the amount of school years that students had been experienced. UNESCO's standards, *International Standards Classification of Education 2011* present an international classification for school levels (ISCED, 2012);

ISCED Level 0- Early childhood education (0-5 years old)

ISCED Level 1- Primary education

ISCED Level 2- Lower secondary education

ISCED Level 3- Upper secondary education

ISCED Level 4- Post secondary education (Non-territory or territory education)

ISCED Level 5- Short cycle

ISCED Level 6- Bachelor's or equivalent level

ISCED level 7- Master's or equivalent level

ISCED level 8- Doctoral or equivalent level

According to this classification TIMSS administrations 4^{th} and 8^{th} grade target groups are described as:

Fourth-grade: All students enrolled in the grade that represents four years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 9.5 years.

Eighth grade: All students enrolled in the grade that represents eight years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 13.5 years. (LaRoche, Joncas & Foy, 2016, p.3.3)

In this research, target populations are about eighth grade. In Table 1, attendant countries are given together with numbers of schools, numbers of classrooms and numbers of students at eighth grade.

Table 1. Numbers of Schools, Classrooms and Students Samples for TIMSS 2015-Eighth Grade

Country	Schools	Classrooms	Students
Australia	285	941	10,338
Bahrain	105	201	4,918
Botswana	159	169	5,964
Canada	276	409	8,757
Chile	171	173	4,849
Chinese Taipei	190	216	5,711
Egypt	211	215	7,822
England	143	215	4,814
Georgia	153	188	4,035
Hong Kong SAR	133	175	4,155
Hungary	144	278	4,893
Iran, Islamic Rep. of	250	251	6,130
Ireland	149	526	4,704
Israel	200	603	5,512
Italy	161	230	4,481
Japan	147	231	4,745
Jordan	252	260	7,865
Kazakhstan	172	239	4,887
Korea, Rep. of	150	317	5,309
Kuwait	168	191	4,503
Lebanon	138	185	3,873
Lithuania	208	269	4,347
Malaysia	207	326	9,726
Malta	48	224	3,817
Morocco	345	373	13,035
New Zealand	145	489	8,142
Norway	143	239	4,697
Oman	301	356	8,883
Qatar	131	250	5,403
Russian Federation	204	226	4,780
Saudi Arabia	143	149	3,759
Singapore	167	334	6,116
Slovenia	148	467	4,257
South Africa	292	334	12,514
Sweden	150	206	4,090
Thailand	204	205	6,482
Turkey	218	220	6,079
United Arab Emirates	477	746	18,012
United States	246	429	10,221
Total	**7,534**	**12,055**	**252,625**

Note: Derived from TIMSS 2015 Assessment Framework.

In TIMSS 2015 administration, 39 countries had been attended to the study at 8^{th} grade level (except the benchmarking countries). The benchmarking seven countries are not taken into this research's extent. Three considered levels of this study are student, school and country levels.

All the national sampling process for TIMSS 2015 had been done under a plan by TIMSS sampling experts and the attendant countries' national research coordinators. The sample groups of countries for TIMSS administrations are decided as a two-stage stratified sampling. First clusters are appeared as schools and second clusters as classes within schools. Sampling is done with stratum in order to provide more reliable sampling and increase the representativeness of each feature school in a country (Joncas & Foy, n.d.; LaRoche, Joncas & Foy, 2016).

At first sampling stage, schools are chosen according to a stratification method. This method might be implicit or explicit. Picking same numbers of schools from each region of a country might be an example for explicit stratification. On the other hand, within each explicit strata, sorting the schools according to socio-economic status, low-medium-high, might be an example for implicit stratification. Additionally, at TIMSS administrations school size is always reckoned with as an implicit stratification variable in order to make a systematic random selection up to the probability of school size. School size is decided mostly as the up-to-date number of students in the target grade. At second sampling stage, classes are incorporated to the study as intact. Class sampling is done by Within-School Sampling Software (WinW3S) generated by IEA Data Processing and Research Center (DPC) and Statistics Canada (LaRoche, Joncas & Foy, 2016).

In the current study, it is aimed to take all 39 countries (except benchmarking countries) including 7,534 schools and 252,625 students. During the analysis, Egypt was excluded because of missing data at country-level, and Israel and Russian Federation were excluded because of list wise missing data at school level. Therefore, after all preliminary analysis had been done; the analysis was conducted with 36 countries, 6,586 schools and 219,788 students.

3.3 Data collection

The data collection process is done by IEA in partnership with each country's National Research Coordinators. Because the data is used from "Data Center" at IEA's webpage TIMSS and PIRLS International Study Center, in this section the data gathering process is summarized as given at that webpage. At each TIMSS administration, similar steps are used with small changes according to the current changes in technology and methodological updates.

The process of administering the instruments and collecting the data is done by each country's National Research Coordinators with the guidance of IEA's operation manuals. The following steps are followed by the National Research Coordinators:

1. Communicating with sampled schools and classes
2. Supervising translation process and getting the assessment instruments ready
3. Conducting the TIMSS 2015 assessments
4. Scoring the constructed response items
5. Forming the data files (Johansone, 2016, p.6.4)

Before these operations three other steps are operated by the TIMSS & PIRLS International Study Center. These operations are "sampling schools, translation and translation verification of the assessment instruments, and layout verification of the assessment instruments".

The process from contacting and tracking the schools till administering the assessment documents are given in Figure 2 as given on Johansone (2016, p.6.5).

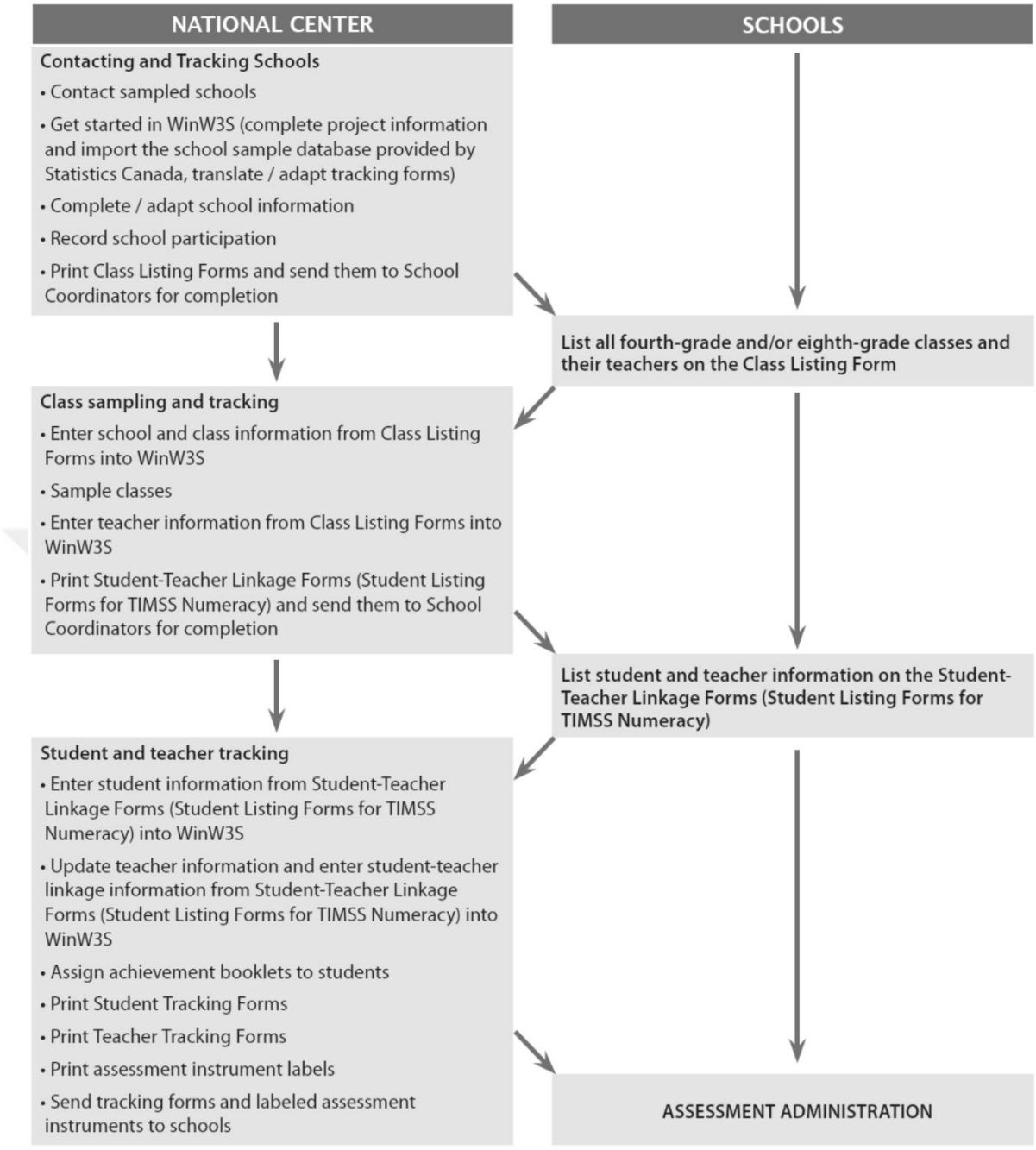


Figure 2. Diagram of the sampling procedures and preparations for the assessment administration implemented by national centers and schools
(Johansone, 2016, p.6.5)

3.4 Instruments

During the current study, the data gathered by mentioned instruments below were used directly. As the outcome variable, students' mathematics scores which were gathered by TIMSS 2015 eighth grade mathematics achievement test were used. As

student level predictor variable, determined variable which was gathered by TIMSS 2015 eighth grade student questionnaire was used. As school level predictor variables, determined variables which were gathered by TIMSS 2015 eighth grade mathematics teacher questionnaire were used. As country level predictor variables, determined variables which were gathered by TIMSS 2015 curriculum questionnaire were used. These instruments are explained briefly at following sections.

3.4.1 Eighth grade TIMSS 2015 mathematics achievement test

TIMSS & PIRLS International Study Center uses item development procedures in order to develop new mathematics items which are also given in Crocker and Algina (2008). The process has the following steps:

1. Updating the frameworks in accordance with the upcoming assessment
2. Generating items and their scoring rubrics
3. Conducting a field-test on a large representative sample
4. Selecting the items in accordance with the frameworks, field-test results, and existing items which had already used at previous cycles
5. Organizing trainings for scoring the open-ended items reliably (Mullis, Drucker, Preuschoff, Arora, Stanco, 2012, p.6; Mullis, Cotter, Fishbein and Centurino, 2016, p.1.2)

TIMSS 2015 mathematics assessment contains number, algebra, geometry, data and chance as content domains. Mathematics content domains can be viewed in detail from http://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap1.pdf at pages 19 to 23. Besides the content domains, tests are formed in accordance with knowing, applying and reasoning as cognitive domains. Mathematics items are given in multiple choice or constructed response items. In order to get reliable scores,

expert people are trained to score the constructed response items (Grønmo, Lindquist, Arora & Mullis, 2016). For TIMSS 2015 eighth grade constructed response mathematics items, average reliability of cross-country scoring is 93% for correctness score agreement (Foy et al., 2016).

In 2015 administration, in total 212 eighth grade mathematics achievement items are used in different booklets. Items are reviewed according to some standards by the help of a signaling system. With this signaling system, item discrimination, item difficulty, Rasch difficulty estimate for each country, response option analysis and scoring reliability are done and according to these statistics, items are reviewed or eliminated from tests. As overall reliability, the median of the Cronbach's Alpha reliability coefficients for all countries are taken. For TIMSS eighth grade mathematics tests the international median Cronbach's alpha reliability is 0.88 (Foy et al., 2016). The reliability coefficients are available per each country in the same document.

For achievement tests, TIMSS uses 500 as the scale center-point and students' performances generally change between 300 and 700. In this current study, the mathematics scales do not have been used directly. Instead, the data collected by those scales are used.

3.4.2 Eighth grade student questionnaire

For all context questionnaires used in TIMSS administrations, similar consecutive steps are followed. TIMSS 2015 context questionnaires development steps for both student and teacher questionnaires are as follows:

1. Updating the context questionnaire framework for 2015
2. Modifying and developing new context questionnaire items by staff at the TIMSS & PIRLS International Study Center
3. Reviewing and revising the questionnaires by the Questionnaire Item Review Committee (QIRC) and NRCs
4. Administering the TIMSS 2015 field test
5. Using the field test results to refine the questionnaires (Hooper, 2016, p.2.1)

Eighth grade student questionnaire utilized in TIMSS 2015 administration can be seen on Section 6, pages 174-196 in *TIMSS 2015 User Guide for International Database-Supplement 1* (https://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UG_Supplement1.pdf). The eighth grade student questionnaire has 26 main items that most of them have sub-items in yes-no or "agree a lot, agree little, disagree a little, disagree a lot" answer formats. The student questionnaire has items about students' home and school routines, home environment, learning environment in school, self-perception and attitudes about mathematics and science lessons.

3.4.3 Eighth grade mathematics teacher questionnaire

Mathematics teacher questionnaire is one of the complementary instruments used at TIMSS administrations in order to support the achievement test results and reveal the latent variables of achievement. In a parallel way with the process of achievement tests, also for contextual questionnaires Cronbach's alpha coefficients per country and international median of these coefficients are calculated. As a criteria, it is expected that Cronbach's alpha coefficient to be above 0.70 (Martin et al., 2016).

The teacher questionnaire utilized in TIMSS 2015 administration can be seen on the Section 7, pages 246-260 in *TIMSS 2015 User Guide for International Database-Supplement 1*.(https://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UG_Supplement1.pdf). The questionnaires have been updated partially administration by administration, and so the questionnaires have been slightly changed compared with previous administrations.

3.4.4 TIMSS curriculum questionnaire

In each cycle, TIMSS collects various data about many different contexts by conducting background questionnaires to students, teachers, school administrations and curriculum experts of countries which affect learning. Curriculum questionnaires form a part in the contextual background questions of TIMSS. They are answered by National Research Coordinator (NRC) of each country. Countries present data about their educational systems and curricula at parts of general, mathematics and science modules (Mullis et al., 2012).

TIMSS curriculum questionnaire provides a crucial source to understand the teaching and learning of mathematics at worldwide level. Curriculum questionnaire in TIMSS 2015 is designed to gather data about both the structure of the education systems of countries and organization, content, and implementation of the mathematics and science curricula of participant countries which aims to support other contextual questionnaires. Curriculum questionnaire exhibits for TIMSS 2015 are given as follows:

1. Years of compulsory schooling
2. National policies on school entry and promotion
3. Types of early childhood/ pre-primary education

4. Topics covered by early childhood/pre-primary education curriculum guidance documents
5. Structural characteristics of curricula
6. Components prescribed by mathematics and science curricula
7. Instructional time devoted to mathematics and science curricula
8. Number of TIMSS topics intended to be taught by the end of fourth and eighth grade
9. Methods used to evaluate the implementation of mathematics and science curricula
10. National policies regarding examinations with consequences for students
11. Processes for approving instructional materials for mathematics and science
12. National policies regarding use of technology in mathematics and science instruction
13. Main preparation routes and current requirements for teachers
14. Main preparation routes and current requirements for principals. (Mullis, Martin, Goh & Cotter, 2016)

The curriculum questionnaire used in TIMSS 2015 administration can be seen on TIMSS 2015 *User Guide for the International Database-Supplement 1* (Foy, 2017). All the related data used in the processes of this thesis are downloaded from http://timssandpirls.bc.edu/timss2015/international-database/ as SPSS or Excel formats.

3.4.5 Variables and the operational definitions

In this part, the variables which are considered in the scope of this research are defined operationally. The variables at school level are aggregated from classroom

level variables which are gathered based on mathematics teacher questionnaire. Furthermore, index variables at both school and country level were converted to standard z scores in order to standardize the deviations.

In this study, in order to determine the variables at student, school and country levels which are related with the mathematics achievement levels of 8th grade students, the data of:

- Eighth grade students' 1st plausible values of mathematics achievements,
- Determined variable of student characteristic from the answers of TIMSS eighth grade student background questionnaire at student level,
- Determined variables of classroom context from the answers of TIMSS eighth grade mathematics teacher questionnaire aggregated at school level,
- Determined variables of curriculum context from the answers of TIMSS curriculum questionnaire by each countries' national research coordinators' at country level

are employed. Abbreviations of variables or index names used during analysis are given at Table 2.

Outcome Variable:

Mathematics achievement scores (MATH): The calculated 1st plausible value of 8th grade students' mathematics scores in TIMSS 2015. Wu (2005) stated that "even one plausible value per student is sufficient to adequately recover population parameters" (p.116). Hence, it can be considered that using only 1st plausible value does not have an inconvenience.

Table 2. The Abbreviations of Variables or Indexes

Variables/Indexes	Variable or Index Name	Abbreviation
Outcome Variables	Students' mathematics achievement scores	MATH
Student Level Predictor Variable (Level-1)	Index of home study support	HSUP
School Level Predictor Variables (Level-2)	Index of percentage of student taught the topic 'Numbers'	NUMB
	Index of percentage of student taught the topic 'Algebra'	ALGB
	Index of percentage of student taught the topic 'Geometry'	GEOM
	Index of percentage of student taught the topic 'Data and Chance'	DATA
	Index of teachers' preparedness to teach mathematics	PREP
	Index of homework use in learning processes	HMWC
	Index of teachers' BS degree	DEGR
	Index of working load of teachers	LOAD
	Number of student in the classes	STUD
	Weekly mathematics class time	TIME
	Homework frequency	HWFR
Country Level Predictor Variables (Level-3)	Index of components prescribed by mathematics curriculum	COMP
	Instructional time devoted to mathematics curriculum	INST
	Number of TIMSS math topics intended to be taught by the end of eighth grade	NTOP
	Index of methods used to evaluate the implementation of mathematics	EVAL
	Mathematics and numeracy skills at pre-primary education	MPPE
	Classroom tracking/grouping by achievement	CGRP

Predictor Variables

1st Level Predictor Variable (Student Level):

Index of home study support (HSUP): This index is derived by IEA based on the question "Do you have any of these things in your home?" which is stated on *Supplement 3* as follows (Foy, 2017, p.23). The answers to the statements: "Your own room" and "Internet connection" and response options 1= "*Yes*" and 2= "*No*", the answers recoded as, 0= "*Neither own room nor internet connection*", 1= "*Either own room or internet connection*" and 2= "*Both own room and internet connection*". Hence, the data of this index variable is interpreted as higher the score of this index variable refers the more home study support the student has.

2nd Level Predictor Variables (School Level):

From eighth grade mathematics teacher questionnaire, teachers' answers for some questions are used to form index variables by IEA. These variables are: index of percentage of students taught the topic 'Numbers', index of percentage of students taught the topic 'Algebra', index of percentage of students taught the topic 'Geometry', index of percentage of students taught the topic 'Data and chance' and index of teachers' BS degree.

1. Index of percentage of students taught the topic 'Numbers' (NUMB): This index is derived by IEA as an answer to the question:

> The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when the students in this class have been taught each topic. If a topic was in the curriculum before the <eighth grade>, please choose "Mostly taught before this year." If a topic was taught half this year but not yet completed, please choose "Mostly taught this year." If a topic is not in the curriculum, please choose "Not yet taught or just introduced. (Foy, 2017, p.33)

This index includes the answers from five statements which refers five objectives about the topic 'Numbers'. First the percent of students whose teachers selected "mostly taught before this year" or "mostly taught this year" are computed, and then average across the percentages of those five objectives are computed and the index is formed. The higher the value is, more students are taught about the topic 'Numbers'.

2. Index of percentage of students taught the topic 'Algebra' (ALGB): This index is derived by IEA as an answer to the same question at previous index. This index includes the answers from six statements which refers six objectives about the

topic ‘Algebra’. First the percent of students whose teachers selected “mostly taught before this year” or “mostly taught this year” are computed, and then average across the percentages of those six objectives are computed and the index is formed.

3. Index of percentage of students taught the topic ‘Geometry’ (GEOM): This index is derived by IEA as an answer to the same question at previous index. This index includes the answers from six statements which refers six objectives about the topic ‘Geometry’. First the percent of students whose teachers selected “mostly taught before this year” or “mostly taught this year” are computed, and then average across the percentages of those six objectives are computed and the index is formed.

4. Index of percentage of students taught the topic ‘Data and Chance’ (DATA): This index is derived by IEA as an answer to the same question at previous index. This index includes the answers from three statements which refers three objectives about the topic ‘Data and Chance’. First the percent of students whose teachers selected “mostly taught before this year” or “mostly taught this year” are computed, and then average across the percentages of those three objectives are computed and the index is formed.

5. Index of teachers’ BS degree (DEGR): This index is derived by IEA as an answer to the question “During your <post secondary>education, what was your major or main area(s) of study?” (Foy, 2017). The answers are coded as a multinomial data. The data of the index variable teachers’ BS degree has been recoded by the aim of interpreting the results easier. Recoding for teachers’ BS degree is given in Table 3.

Table 3. Recoding for the Index of Teachers' BS Degree

Statement	Original Code	New Code
Major in math and math education	1	4
Major in math but not in math education	2	2
Major in math education but not in math	3	3
All other majors	4	1
No formal education beyond upper secondary	5	0
Omitted or invalid	9	9

On the other hand, the data for the following three variables were taken directly from the data set of TIMSS 2015 and aggregated to school level.

6. Number of students in the classes (STUD) is a continuous variable. This variable was formed from the question, "How many students are in the class?". The higher number of students in the classes means a crowded learning environment.

7. Weekly mathematics class time (TIME) is a continuous variable. This variable was formed from the question, "In a typical week, how much time do you spend teaching mathematics to the students in this class?" The higher weekly mathematics class time means that school spares more time on mathematics classes.

8. Homework frequency (HWFR): The variable score ranges from 1 (*never*) to 5 (*every day*) as a multinomial variable. This variable was formed from the question, "How often do you usually assign mathematics homework to the students in this class?" The minimum score means teacher never assign homework to the class and the maximum score means teacher assigns homework to the class every day.

On the other hand, three index variables are formed special for this study as taking the sum of related answers. Those answers were first recoded in order to make the interpretation of the results easier. The related questions from mathematics teacher questionnaire and recodes for answers are given below.

9. Index of teacher preparedness to teach mathematics topics (PREP): In TIMSS 2015 administration, the question "How well prepared do you feel you are to

teach the following topics?" requires 20 sub-items to be answered and as, 1: *Not applicable*, 2: *Very well prepared*, 3: *Somewhat prepared*, 4: *Not well prepared*, 9: *Omitted or invalid*. As an index variable, the total score of the answers to sub-items are calculated after the answers are recoded as Table 4.

Table 4. Recoding for the Teacher Preparedness about Mathematics Topics

Statement	Original Code	New Code
Not applicable	1	9
Very well prepared	2	2
Somewhat prepared	3	1
Not well prepared	4	0
Omitted or invalid	9	9

The index scores were used to range from 0 to 40, which were converted to standard z scores. Lower scores mean the responsible teacher of the class has been prepared to less TIMSS mathematics topics and higher scores mean the responsible teacher of the class has been prepared to more TIMSS mathematics topics.

10. Index of homework use in learning processes (HMWC): In TIMSS 2015 administration, the question "How often do you do the following with the mathematics homework assignments for the TIMSS class?" requires 5 sub-items to be answered as, 1: *Always or almost always*, 2: *Sometimes*, 3: *Never or almost never*, 9: *Omitted or invalid*. The sub-items are as:

> a) Correct assignments and give feedback to students, b) Have students correct their own homework, c) Discuss the homework in class, d) Monitor whether or not the homework was completed, e) Use the homework to contribute towards students' grades or marks (IEA, 2014, p.6)

As an index variable, the total score of the answers to sub-items are calculated after the answers are recoded as Table 5.

Table 5. Recoding for the Homework Use in Learning Processes

Statement	Original Code	New Code
Always or almost always	1	2
Sometimes	2	1
Never or almost never	3	0
Omitted or invalid	9	9

The index scores were used to range from 0 to 10 which were converted to standard z scores. The higher the score of this index means the more use of homework in the learning processes.

11. Index of working load of teacher (LOAD): In TIMSS 2015 administration, the question "Indicate the extent to which you agree or disagree with each of the following statements." requires 8 sub-items to be answered as, 1: *Agree a lot*, 2: *Agree a little*, 3: *Disagree a little*, 4: *Disagree a lot*, 9: *Omitted or invalid*. The sub-items are as:

> a) There are too many students in the classes, b) I have too much material to cover in class, c) I have too many teaching hours, d) I need more time to prepare for class, e) I need more time to assist individual students, f) I feel too much pressure from parents, g) I have difficulty keeping up with all of the changes to the curriculum, h) I have too many administrative tasks (IEA, 2014, p.6)

As an index variable, the total score of the answers to sub-items are taken after the answers are recoded as Table 6. The index scores were used to range from 0 to 24, which were converted to standard z scores. The higher the score of this index means the busier the teacher is.

Table 6. Recoding for the Working Load of Teacher

Statement	Original Code	New Code
Agree a lot	1	3
Agree a little	2	2
Disagree a little	3	1
Disagree a lot	4	0
Omitted or invalid	9	9

3rd Level Predictor Variables (Country Level):

The questions and variables used from the curriculum questionnaire for the current study are as follows:

1. Index of components prescribed by mathematics curriculum (COMP): In TIMSS 2015 administration, the question "What does the mathematics curriculum prescribe?" requires 4 sub-items to be answered as "*Yes/No*". The index variable was formed as taking the sum of the answers as coding 1=*Yes* and 0=*No*. This variable includes an index point formed by the answers of whether goals and objectives, instructional methods and processes, materials and assessment methods or activities are stated in the national curriculum or not. The index score ranges from 0 to 4. Higher score indicates more items are accurate in the curriculum and this shows a more centralized education system whereas low total score refers more freedom on educational processes.

2. Instructional time devoted to mathematics curriculum (INST): This variable is calculated as "principal reports of school days per year" times "teacher reports of weekly mathematics instructional hours" divided by "principal reports of school days per week". Hence, this variable is given in terms of the amount of hours. Higher score means that country use more time on mathematics classes.

The answers from the curriculum data are not taken for this variable because of high proportion of missing data. So, the data for this variable is formed by a formula by TIMSS and PIRLS Study Center;

$$\begin{pmatrix}\text{Hours per Year}\\ \text{for Mathematics}\\ \text{Instruction}\end{pmatrix} = \frac{\begin{matrix}\text{Teacher Reports of}\\ \text{Weekly Mathematics}\\ \text{Instructional Hours}\end{matrix}}{\begin{matrix}\text{Principal Reports of}\\ \text{School Days per Week}\end{matrix}} \times \begin{pmatrix}\text{Principle Reports of}\\ \text{School Days per Year}\end{pmatrix} \qquad (3.1)$$

3. Number of TIMSS topics intended to be taught by the end of eighth grade (NTOP): This variable includes the answer by attendee countries' NRCs to how many topics out of 20 have been taught to all or almost all students till to TIMSS 2015 administration. This variable ranges between 0 and 20. Max score means, all TIMSS topics are covered in the country's curriculum.

4. Index of methods used to evaluate the implementation of mathematics curriculum (EVAL): This variable includes the total point of answers by attendee countries' NRCs to the question, "How is the mathematics curriculum implementation evaluated?", whether the evaluations are done as visits by inspectors, research programs, school self-evaluation, national or regional examinations. The index score ranges from 0 to 4. Higher score indicates that evaluation of the curriculum is considered important whereas low total score refers less care about evaluation of the curriculum implementation.

5. Mathematics and numeracy skills at the pre-primary education (MPPE): If the pre-primary education program includes mathematics and numeracy skills, than this variable takes 1 otherwise it takes 0.

6. Classroom tracking/grouping by achievement (CGRP): If there is an assigning system for students in the country, this variable takes 1 otherwise it takes

zero. The operational definitions for curriculum level variables are done in parallel with the explanations in Mullis, Martin, Goh and Cotter (2016).

3.5 Preparation of data

TIMSS 2015 data were downloaded from the related database. In order to use the data in HLM, the level-1 file which includes the variables at student level, level-2 file which includes the variables at school level, level-3 file which includes the variables at country level were formed. In order to obtain the school level data, data gathered from teachers were aggregated to each school by SPSS 23.

In order to prepare the data for operating the analysis; the data were examined in detail according to the given ID structure for schools, classes, students and teachers. The structures of IDs are given on Table 7 (Johansone, 2016, p.6.9).

Table 7. Hierarchical Identification System Codes Used to Link Schools, Classes, Students and Teachers

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	0001
Class	School+Class within the school	CCCCKK	000101 000102
Student	School+Class within the school+Student within the class	CCCCKKSS	00010101 00010201
Teacher	School+Teacher within the school+Linkage number to the sampled class	CCCCTTLL	00010101 00010201

As seen on Table 7, teacher identification numbers have four digits of country identification number, which is followed by two digits of teacher number. At the end, there is another two digit classroom linkage number to a class is needed to identify if a teacher teaches a part of a class or whole of it. The data will be prepared as considering these codes in order to prevent teacher-classroom confusion.

In TIMSS 2015 administration, data shows a strictly hierarchical structure as with schools nested in countries, classes nested within schools and students nested within classes (Johansone, 2016). On the other hand, teacher IDs have a problem about matching with classes and students. Especially in some countries (e.g.Australia, Canada, Georgia) one student has more than one teacher, and because for classroom context the data of the answers of teacher questionnaire would be used, that caused problem to prepare data to use at three-level HLM analysis. The problem based on the fact that a lower level member, in this case student, can not be assigned more than one higher level class, in this case teacher. Without aggregation, utilizing HLM would not be possible for the current data set. Therefore, the data at classroom level which gathered by teacher questionnaires had to be aggregated and used at school-level.

3.6 Data analysis

The purpose of this study is to determine if Trends in International Mathematics and Science Study (TIMSS) 2015 eighth grade mathematics achievement scores vary between countries and if there is a variation, determine the curriculum, classroom and student related factors behind that variation. It is also aimed to consider the effect of interactions in different level variables to reveal conditional circumstances For this aim, hierarchical linear modeling (HLM) is utilized to analyze the related data. Data is firstly predisposed by SPSS23.0 to MDM file format by Hierarchical Linear & Nonlinear Modeling software HLM 7. Then the main analyses had been done by HLM7 (Raudenbush, Bryk, & Congdon, 2011).

3.6.1 The hierarchical structure of the data

In social organizations, people interact with their environments as it is with their work place, school or the country. This interaction between levels forms a two way influence. The units at one level can affect the units at another level and vice versa. For such organizations, nature of the data should be considered during research processes and analyses need to be done according to that nature. Hierarchical structure can be seen in many different instances, such as:

- Cross national studies which individuals are nested within their national units,
- Organizational studies where individuals are nested within departments,
- Family studies which family members are nested within families,
- Methodological studies where respondents are nested within interviewers,
- Longitudinal studies where distinct observations are nested within individuals,
- Meta-analysis where subjects are nested within different studies. (Hox, 2002, p.1)

If these hierarchical structures are classified as lower and higher level variables, for cross national studies, individuals refer the lower level variable whereas national units refer the higher level variable. Special to educational areas, it is seen that the data are nested one in each other as students are nested in classrooms, classrooms are nested in schools, schools are nested in districts and districts are nested in counties. For studies done in educational area, the hierarchical structure of the data has to be taken into consideration (Hox, 2002).

For hierarchical structure data, there are three options for researchers to work with the data, namely aggregation, disaggregation and hierarchical linear modeling (Hofmann, 1997).

3.6.2 Aggregation and disaggregation

Aggregation represents the process of moving a lower level variable to higher level and disaggregation as quite the opposite implies the process of moving a higher level variable to lower level (Hox, 2002; Woltman, Feldstain, Mackay & Rocchi, 2012). To make it clear, aggregation aims to unite the small units together and get a bigger unit, such as getting the average height of all students in a class and use it as a class variable. On the other hand, disaggregation aims to distribute a characteristic which belong to a bigger unit and get smaller units, such as giving the class average point to each student and representing students with that point.

For hierarchical structure data, both aggregation and disaggregation have some handicaps. As applying aggregation, lower level individual differences are ignored, which means individual variability is lost from the data. Aggregation causes to detachment of individual and group effects from the outcome variable. On the other hand, disaggregation ignores the presence of group differences. As applying disaggregation the assumption of independence of errors are contravened (Woltman et al., 2012). Briefly, aggregated model ignores possibly meaningful individual variances and disaggregated model contravenes the statistical assumptions and considers the influence of classes on the number of students. With the knowledge of these weaknesses of both aggregation and disaggregation, hierarchical linear modeling designed in order to overwhelm the before mentioned deficiencies (Hofmann, 1997).

3.6.3 Hierarchical linear models

Hierarchical linear models (HLM) have some advantages which cannot be ignored compared to both aggregation and disaggregation. First of all, HLM clearly perceive

that “individuals within a particular group may be more similar to one another than individuals in other groups” (Hofmann, 1997, p.726) which means that these models realize the partial interdependence of individuals which nested in the same group. Secondly, HLM makes it possible to apply both the group and individual predictors in the levels they belong to and interrogate both lower and higher level variances on outcome. In another word, as applying individual independent variables at individual level and group independent variables at group level, both individual and group variables can be modeled in individual outcomes (Hofmann, 1997; Osborn, 2000). In addition, HLM presents more convenient error terms that control possible adhesion arising nesting impressions and states more correct error terms (Newman, Newman & Salzman, 2010).

Although HLM allow analyzing multiple levels of data, in this research three-level HLM is considered with a student characteristic at level-1, classroom characteristics as aggregated to school at level-2 and country characteristics at level-3 variables, where the level-1 units are nested within level-2 units and level-2 unites are nested within level-3 units. In three-level model, there are i=1, …, n_{jk} level-1 units (e.g., students) which are nested in each of j=1, …J_k level-2 units (e.g. schools, which are nested within each of k=1, …, K level-3 units (e.g. countries). Three-level model includes three sub-models. Model formulations are given at following part.

Model Formulation

For level-1, level-2 and level-3 variables, the model equations can be given as follows (Raudenbush, Bryk, Cheung, Congdon, and du Toit, 2011, pp.63-66).

Level-1: $Y_{ijk} = \pi_{0jk} + \pi_{1jk}a_{1jk} + \pi_{2jk}a_{2jk} + \cdots + \pi_{pjk}a_{pjk} + e_{ijk}$

$$= \pi_{0jk} + \sum_{p=1}^{p} \pi_{pjk}a_{pjk} + e_{ijk} \quad (3.2)$$

where

Y_{ijk} : outcome measure for level-1 student *i* in level-2 school *j* and level-3 country k,

π_{pjk} (p=0,1, …, P) are level-1 coefficients,

π_{0jk} is a level-1 intercept and π_{1jk},…, π_{pjk} are level-1 slopes

a_{pjk} is a level-1 predictor p for student i in level-2 school j and level-3 country k,

e_{ijk} is the level-1 random effect,

σ^2 is the variance of e_{ijk}, which is the variance of level-1.

It is assumed that $e_{ijk} \sim N(0, \sigma^2)$. In level-2 model, each level-1 coefficients are become an outcome variable.

Level-2: $\pi_{pjk} = \beta_{p0k} + \beta_{p1k}X_{1jk} + \beta_{p2k}X_{2jk} + \cdots + \beta_{pQ_pk}X_{Q_pjk} + r_{pjk}$

$$= \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk}X_{qjk} + r_{pjk} \quad (3.3)$$

where

β_{pqk} (q= 0, 1, …, Q_p) are level-2 coefficients,

β_{p0k} is a level-2 intercept and $\beta_{p2k} + \cdots + \beta_{pQ_pk}$ are level-2 slopes

X_{qjk} is a level-2 predictor, and

r_{pjk} is a level-2 random effect.

Each level-1 coefficient can be formulated in the model at level 2 as one of three given general forms, as follows (Raudenbush et. al., 2011, p. 65).

1. A level-1 coefficient that is fixed for all level-2 units; e.g.,

$$\pi_{pjk} = \beta_{p0k} \quad (3.4)$$

2. A level-1 coefficient that varies non-randomly among level-2 units, e.g.,

$$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} \quad (3.5)$$

3. A level-1 coefficient that varies randomly among level-2 units, e.g.,

$$\pi_{pjk} = \beta_{p0k} + r_{pjk} \quad \text{or} \quad \pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk} \quad (3.6)$$

In level-3 model, each level-2 coefficients are become an outcome variable.

$$\text{Level-3: } \beta_{pqk} = \gamma_{pq0} + \gamma_{pq1} W_{1k} + \gamma_{pq2} W_{2k} + \cdots + \gamma_{pqS_{pq}} W_{S_{pq}k} + u_{pqk}$$

$$= \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqS} W_{Sk} + u_{pqk} \quad (3.7)$$

where

γ_{pqs} (s= 0, 1, …, S_{pq}) are country level coefficients,

W_{sk} is a level-3 country predictor, and

U_{pqk} is a level-3 country random effect

Each level-2 coefficient can be formulated in the model at level 3 as one of three given general forms, as follows (Raudenbush et. al., 2011, p. 66).

1. A level-2 coefficient that is fixed for all level-3 units; e.g.,

$$\beta_{pqk} = \gamma_{pq0} \quad (3.8)$$

2. A level-2 coefficient that varies non-randomly among level-3 units, e.g.,

$$\beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} \quad (3.9)$$

3. A level-2 coefficient that varies randomly among level-3 units, e.g.,

$$\beta_{pqk} = \gamma_{pq0} + u_{pqk} \quad \text{or} \quad \beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} + u_{pqk} \quad (3.10)$$

With the following sub-titles, six models which had been used during the current study in order to test the research questions are briefly explained. On Figure 3, the flow chart for the research questions based on the models developed during the study can be seen. One of the models has four modes, starting with the

unconditional model ends with the model which has student, school and country level predictors all together. On the other hand, the other models include only school level predictors and only country level predictors in order to obtain the differences in variances with the inclusion of different level predictors.

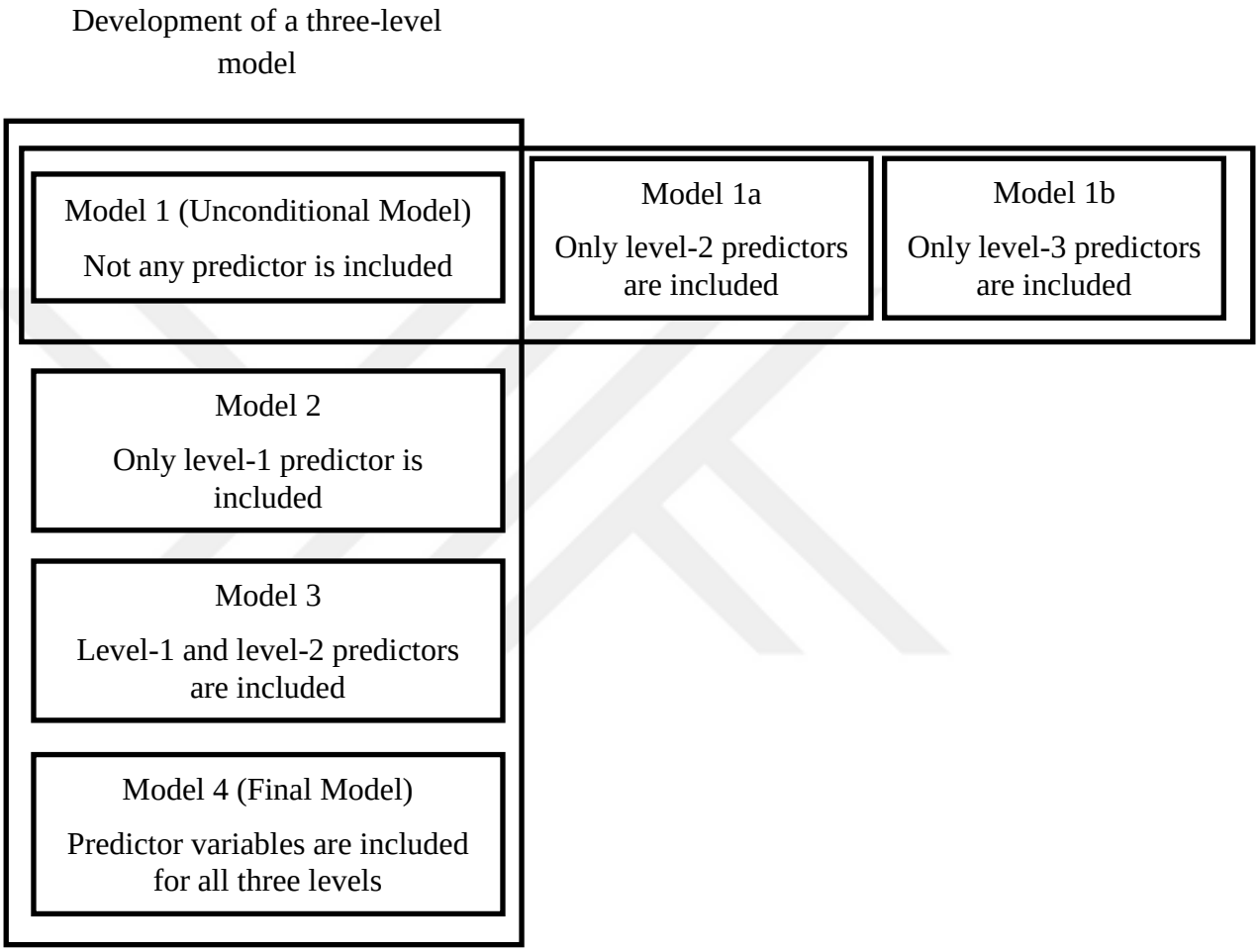


Figure 3. Development scheme of the models

3.6.3.1 Model 1 (The unconditional model)

The unconditional model is formulated as taking students' mathematics achievement scores as an outcome with no predictor variable in the model. This model is mostly used for separating "variation in intercepts and growth rates into components" (Raudenbush et al., 2011, p.82) that lie within and between countries. On the other

hand, this model does not give any information how student, classroom or country characteristics are related.

Outcome variable is predicted via a three-level model. This model is a prerequisite model which needs to be tested in order to operate subsequent models. This basic model is used to determine the source of the variance on dependent variable in terms of level-1, level-2 and level-3 variables. For the current study, the equations for the unconditional model are as follows (Raudenbush et al., 2011; Subedi, 2005).

Level 1: $MATH_{ijk} = \pi_{0jk} + e_{ijk}$ (3.11)

Level2: $\pi_{0jk} = \beta_{00k} + r_{0jk}$ (3.12)

Level 3: $\beta_{00k} = \gamma_{000} + u_{00k}$ (3.13)

Combined Model: $MATH_{ijk} = \gamma_{000} + e_{ijk} + r_{0jk} + u_{00k}$ (3.14)

As seen on composed model, dependent variable is explained only by third stage intercept, and level-1, level-2 and level-3 residuals. For next two models, exploratory analyses were carried out by HLM 7 in order to select the probable predictors for conditional level-2 and conditional level-3 models.

3.6.3.2 Model 1a (Specifying only school level predictors)

In this model, level-2 is formulated as conditional and level-1 and level-3 are formulated as unconditional. Model 1a is established especially with the aim of investigating the variance difference school level predictor variables. At the beginning, it is planned to take eleven predictor variables which were estimated as important predictors for mathematics achievement scores, whereas after the exploratory analysis, it is decided to take eighth of these level-2 predictors into the Model 1a, namely percentage of student taught the topic ‘Numbers’, percentage of

student taught the topic ‘Algebra, teachers’ preparedness to teach mathematics, homework use in learning processes, teachers’ BS degree, working load of teachers, number of students in the classes and homework frequency.

For the Model 1a, conditional level-2 model and unconditional level-1 and level-3 models are formulated in this study as follows.

Level 1: $MATH_{ijk} = \pi_{0jk} + e_{ijk}$

Level 2: $\pi_{0jk} = \beta_{00k} + \beta_{01k}*(HWFR_{jk}) + \beta_{02k}*(STUD_{jk}) + \beta_{03k}*(NUMB_{jk})$

$+ \beta_{04k}*(ALGB_{jk}) + \beta_{05k}*(PREP_{jk}) + + \beta_{06k}*(HMWC_{jk}) + \beta_{07k}*(LOAD_{jk})$

$+ \beta_{08k}*(DEGR_{jk}) + r_{0jk}$ (3.15)

Level 3: $\beta_{00k} = \gamma_{000} + u_{00k}$

$\beta_{01k} = \gamma_{010} + u_{01k}$

$\beta_{02k} = \gamma_{020} + u_{02k}$

$\beta_{03k} = \gamma_{030} + u_{03k}$

$\beta_{04k} = \gamma_{040} + u_{04k}$

$\beta_{05k} = \gamma_{050} + u_{05k}$

$\beta_{06k} = \gamma_{060} + u_{06k}$

$\beta_{07k} = \gamma_{070} + u_{07k}$

$\beta_{08k} = \gamma_{080} + u_{08k}$ (3.16)

Combined Model for Model 1a:

$MATH_{ijk} = \gamma_{000} + \gamma_{010}*HWFR_{jk} + \gamma_{020}*STUD_{jk} + \gamma_{030}*NUMB_{jk}$

$+ \gamma_{040}*ALGB_{jk} + \gamma_{050}*PREP_{jk} + \gamma_{060}*HMWC_{jk} + \gamma_{070}*LOAD_{jk}$

$+\gamma_{080}*DEGR_{jk} + r_{0jk} + u_{00k} + u_{01k}*HWFR_{jk} + u_{02k}*STUD_{jk} + u_{03k}*NUMB_{jk}$

$+ u_{04k}*ALGB_{jk} + u_{05k}*PREP_{jk} + u_{06k}*HMWC_{jk} + u_{07k}*LOAD_{jk}$

$+u_{08k}*DEGR_{jk} + e_{ijk}$ (3.17)

3.6.3.3 Model 1b (Specifying only country level predictors)

In this model, level-3 is formulated as conditional and level-1 and level-2 are formulated as unconditional. Because the focus of the current study is the effects and interactions of country level curriculum variables, Model 1b is established. This model is used to figure out if there is a variance difference between the final model and the model which includes only the country level predictors. At the beginning, it is planned to take six predictor variables which were estimated as important predictors for mathematics achievement scores, whereas after the exploratory analysis, it is decided to take only two level-3 predictors into the Model 1b, namely components prescribed by mathematics curriculum and instructional time devoted to mathematics curriculum.

For the Model 1b, conditional level-3 model and unconditional level-1 and level-2 models are formulated in this study as follows.

Level 1: $MATH_{ijk} = \pi_{0jk} + e_{ijk}$

Level 2: $\pi_{0jk} = \beta_{00k} + r_{0jk}$

Level 3: $\beta_{00k} = \gamma_{000} + \gamma_{001}(INST_k) + \gamma_{002}(COMP_k) + u_{00k}$ (3.18)

Combined Model for Model 1b:

$MATH_{ijk} = \gamma_{000} + \gamma_{001}(INST_k) + \gamma_{002}(COMP_k) + u_{00k} + r_{0jk} + e_{ijk}$ (3.19)

3.6.3.4 Model 2 (Specifying student-level predictor variables)

In this model, level-1 is formulated as conditional and level-2 and level-3 are formulated as unconditional. One multinomial predictor variable which is related with students' socio-economic status is selected based on the literature and significant correlation between this level-1 predictor variable, namely home study support, and outcome variable, mathematics achievement scores.

For the model 2, conditional level-1 model and unconditional level-2 and level-3 models are formulated in this study as follows.

Level 1: $MATH_{ijk} = \pi_{0jk} + \pi_{1jk} * (HSUP_{ijk}) + e_{ijk}$ (3.20)

Level 2: $\pi_{0jk} = \beta_{00k} + r_{0jk}$

$\pi_{1jk} = \beta_{10k} + r_{1jk}$

Level 3: $\beta_{00k} = \gamma_{000} + u_{00k}$

$\beta_{10k} = \gamma_{100} + u_{10k}$

Combined Model for Model 2:

$MATH_{ijk} = \gamma_{000} + \gamma_{100} * (HSUP_{ijk}) + r_{1jk} * (HSUP_{ijk}) + u_{10k} * (HSUP_{ijk}) + u_{00k} + r_{0jk} + e_{ijk}$ (3.21)

It is considerable to specify the next level predictor variables and develop the conditional model for level-1 and level-2, unconditional model for level-3, for the variables which have significant parameter variance in school-level equations.

3.6.3.5 Model 3 (Specifying school-level predictor variables)

In this model, level-2 is formulated as conditional additionally to level-1, and level-3 is formulated as unconditional. At the beginning, it is planned to include 11 predictor variables which are formed by classroom characteristics at school level, but after the exploratory analysis, as in the model 1a eighth level-2 predictor variables are considered in the model 3. These variables are homework frequency and teachers' BS degree which are multinomial variables, percentage of student taught the topic 'Numbers', percentage of student taught the topic 'Algebra, teachers' preparedness to teach mathematics, homework use in learning processes, number of students in classes and working load of teachers which are continuous variables.

For the model 3, conditional level-1 and level-2 models and unconditional level-3 model are formulated in this study as follows.

Level 1: $MATH_{ijk} = \pi_{0jk} + \pi_{1jk} *(HSUP_{ijk}) + e_{ijk}$

Level 2: $\pi_{0jk} = \beta_{00k} + \beta_{01k}*(HWFR_{jk}) + \beta_{02k}*(STUD_{jk}) + \beta_{03k}*(NUMB_{jk})$

$+ \beta_{04k}*(ALGB_{jk}) + \beta_{05k}*(PREP_{jk}) + \beta_{06k}*(HMWC_{jk}) + \beta_{07k}*(LOAD_{jk})$

$+\beta_{08k}*(DEGR_{jk}) + r_{0jk}$

$\pi_{1jk} = \beta_{10k} + \beta_{11k}*(HWFR_{jk}) + \beta_{12k}*(STUD_{jk}) + \beta_{13k}*(NUMB_{jk})$

$+\beta_{14k}*(ALGB_{jk}) + \beta_{15k}*(PREP_{jk}) + \beta_{16k}*(HMWC_{jk}) + \beta_{17k}*(LOAD_{jk})$

$+\beta_{18k}*(DEGR_{jk})$ (3.22)

Level 3: $\beta_{00k} = \gamma_{000} + u_{00k}$

$\beta_{01k} = \gamma_{010} + u_{01k}$

$\beta_{02k} = \gamma_{020} + u_{02k}$

$\beta_{03k} = \gamma_{030} + u_{03k}$

$\beta_{04k} = \gamma_{040} + u_{04k}$

$\beta_{05k} = \gamma_{050} + u_{05k}$

$\beta_{06k} = \gamma_{060} + u_{06k}$

$\beta_{07k} = \gamma_{070} + u_{07k}$

$\beta_{08k} = \gamma_{080} + u_{08k}$

$\beta_{10k} = \gamma_{100} + u_{10k}$

$\beta_{11k} = \gamma_{110} + u_{11k}$

$\beta_{12k} = \gamma_{120} + u_{12k}$

$\beta_{13k} = \gamma_{130} + u_{13k}$

$\beta_{14k} = \gamma_{140} + u_{14k}$

$\beta_{15k} = \gamma_{150} + u_{15k}$

$\beta_{16k} = \gamma_{160} + u_{16k}$

$$\beta_{17k} = \gamma_{170} + u_{17k}$$

$$\beta_{18k} = \gamma_{180} + u_{18k} \quad (3.23)$$

Combined Model for Model 3:

$$\begin{aligned}MATH_{ijk} = {} & \gamma_{000} + \gamma_{010}*HWFR_{jk} + \gamma_{020}*STUD_{jk} + \gamma_{030}*NUMB_{jk} + \gamma_{040}*ALGB_{jk} \\ & + \gamma_{050}*PREP_{jk} + \gamma_{060}*HMWC_{jk} + \gamma_{070}*LOAD_{jk} + \gamma_{080}*DEGR_{jk} \\ & + \gamma_{100}*HSUP_{ijk} + \gamma_{110}*HSUP_{ijk}*HWFR_{jk} + \gamma_{120}*HSUP_{ijk}*STUD_{jk} \\ & + \gamma_{130}*HSUP_{ijk}*NUMB_{jk} + \gamma_{140}*HSUP_{ijk}*ALGB_{jk} \\ & + \gamma_{150}*HSUP_{ijk}*PREP_{jk} + \gamma_{160}*HSUP_{ijk}*HMWC_{jk} \\ & + \gamma_{170}*HSUP_{ijk}*LOAD_{jk} + \gamma_{180}*HSUP_{ijk}*DEGR_{jk} + r_{0jk} + u_{00k} \\ & + u_{01k}*HWFR_{jk} + u_{02k}*STUD_{jk} + u_{03k}*NUMB_{jk} + u_{04k}*ALGB_{jk} \\ & + u_{05k}*PREP_{jk} + u_{06k}*HMWC_{jk} + u_{07k}*LOAD_{jk} + u_{08k}*DEGR_{jk} \\ & + u_{10k}*HSUP_{ijk} + u_{11k}*HSUP_{ijk}*HWFR_{jk} + u_{12k}*HSUP_{ijk}*STUD_{jk} \\ & + u_{13k}*HSUP_{ijk}*NUMB_{jk} + u_{14k}*HSUP_{ijk}*ALGB_{jk} \\ & + u_{15k}*HSUP_{ijk}*PREP_{jk} + u_{16k}*HSUP_{ijk}*HMWC_{jk} \\ & + u_{17k}*HSUP_{ijk}*LOAD_{jk} + u_{18k}*HSUP_{ijk}*DEGR_{jk} + e_{ijk} \end{aligned} \quad (3.24)$$

At next and final model, the country level predictor variables, which were included also into the Model 1a at section 3.6.3.2 is added to the model 3 and final model is developed.

3.6.3.6 Model 4 (Specifying country-level predictor variables)

In this model, level-3 is formulated as conditional additionally to level-1 and level-2. The variables components prescribed by mathematics curriculum and instructional time which are continuous variables are added to the model at level-3.

According to the predictor variables of each level which have effect on mathematics achievement, the equations are formed for the fourth model. For the

model 4, conditional level-1, level-2 and level-3 models are formulated in this study as follows.

Level-1: $MATH_{ijk} = \pi_{0jk} + \pi_{1jk}*(HSUP_{ijk}) + e_{ijk}$

Level-2: $\pi_{0jk} = \beta_{00k} + \beta_{01k}*(HWFR_{jk}) + \beta_{02k}*(STUD_{jk}) + \beta_{03k}*(NUMB_{jk})$

$+\beta_{04k}*(ALGB_{jk}) + \beta_{05k}*(PREP_{jk}) + \beta_{06k}*(HMWC_{jk}) + \beta_{07k}*(LOADjk)$

$+\beta_{08k}*(DEGR_{jk}) + r_{0jk}$

$\pi_{1jk} = \beta_{10k} + \beta_{11k}*(HWFR_{jk}) + \beta_{12k}*(STUD_{jk}) + \beta_{13k}*(NUMB_{jk})$

$+\beta_{14k}*(ALGB_{jk}) + \beta_{15k}*(PREP_{jk}) + \beta_{16k}*(HMWC_{jk}) + \beta_{17k}*(LOAD_{jk})$

$+\beta_{18k}*(DEGR_{jk}) + r_{1jk}$

Level-3: $\beta_{00k} = \gamma_{000} + \gamma_{001}(INST_k) + \gamma_{002}(COMP_k) + u_{00k}$

$\beta_{01k} = \gamma_{010} + \gamma_{011}(INST_k) + \gamma_{012}(COMP_k) + u_{01k}$

$\beta_{02k} = \gamma_{020} + \gamma_{021}(INST_k) + \gamma_{022}(COMP_k) + u_{02k}$

$\beta_{03k} = \gamma_{030} + \gamma_{031}(INST_k) + \gamma_{032}(COMP_k) + u_{03k}$

$\beta_{04k} = \gamma_{040} + \gamma_{041}(INST_k) + \gamma_{042}(COMP_k) + u_{04k}$

$\beta_{05k} = \gamma_{050} + \gamma_{051}(INST_k) + \gamma_{052}(COMP_k) + u_{05k}$

$\beta_{06k} = \gamma_{060} + \gamma_{061}(INST_k) + \gamma_{062}(COMP_k) + u_{06k}$

$\beta_{07k} = \gamma_{070} + \gamma_{071}(INST_k) + \gamma_{072}(COMP_k) + u_{07k}$

$\beta_{08k} = \gamma_{080} + \gamma_{081}(INST_k) + \gamma_{082}(COMP_k) + u_{08k}$

$\beta_{10k} = \gamma_{100} + \gamma_{101}(INST_k) + \gamma_{102}(COMP_k) + u_{10k}$

$\beta_{11k} = \gamma_{110} + \gamma_{111}(INST_k) + \gamma_{112}(COMP_k) + u_{11k}$

$\beta_{12k} = \gamma_{120} + \gamma_{121}(INST_k) + \gamma_{122}(COMP_k) + u_{12k}$

$\beta_{13k} = \gamma_{130} + \gamma_{131}(INST_k) + \gamma_{132}(COMP_k) + u_{13k}$

$\beta_{14k} = \gamma_{140} + \gamma_{141}(INST_k) + \gamma_{142}(COMP_k) + u_{14k}$

$\beta_{15k} = \gamma_{150} + \gamma_{151}(INST_k) + \gamma_{152}(COMP_k) + u_{15k}$

$\beta_{16k} = \gamma_{160} + \gamma_{161}(INST_k) + \gamma_{162}(COMP_k) + u_{16k}$

$$\beta_{17k} = \gamma_{170} + \gamma_{171}(INST_k) + \gamma_{172}(COMP_k) + u_{17k}$$

$$\beta_{18k} = \gamma_{180} + \gamma_{181}(INST_k) + \gamma_{182}(COMP_k) + u_{18k} \quad (3.25)$$

Combined Model for Model 4:

$$\begin{aligned}
MATH_{ijk} = {} & \gamma_{000} + \gamma_{001}*INST_k + \gamma_{002}*COMP_k + \gamma_{010}*HWFR_{jk} + \gamma_{011}*HWFR_{jk}*INST_k \\
& + \gamma_{012}*HWFR_{jk}*COMP_k + \gamma_{020}*STUD_{jk} + \gamma_{021}*STUD_{jk}*INST_k \\
& + \gamma_{022}*STUD_{jk}*COMP_k + \gamma_{030}*NUMB_{jk} + \gamma_{031}*NUMB_{jk}*INST_k \\
& + \gamma_{032}*NUMB_{jk}*COMP_k + \gamma_{040}*ALGB_{jk} + \gamma_{041}*ALGB_{jk}*INST_k \\
& + \gamma_{042}*ALGB_{jk}*COMP_k + \gamma_{050}*PREP_{jk} + \gamma_{051}*PREP_{jk}*INST_k \\
& + \gamma_{052}*PREP_{jk}*COMP_k + \gamma_{060}*HMWC_{jk} + \gamma_{061}*HMWC_{jk}*INST_k \\
& + \gamma_{062}*HMWC_{jk}*COMP_k + \gamma_{070}*LOAD_{jk} + \gamma_{071}*LOAD_{jk}*INST_k \\
& + \gamma_{072}*LOAD_{jk}*COMP_k + \gamma_{080}*DEGR_{jk} + \gamma_{081}*DEGR_{jk}*INST_k \\
& + \gamma_{082}*DEGR_{jk}*COMP_k + \gamma_{100}*HSUP_{ijk} + \gamma_{101}*HSUP_{ijk}*INST_k \\
& + \gamma_{102}*HSUP_{ijk}*COMP_k + \gamma_{110}*HSUP_{ijk}*HWFR_{jk} \\
& + \gamma_{111}*HSUP_{ijk}*HWFR_{jk}*INST_k + \gamma_{112}*HSUP_{ijk}*HWFR_{jk}*COMP_k \\
& + \gamma_{120}*HSUP_{ijk}*STUD_{jk} + \gamma_{121}*HSUP_{ijk}*STUD_{jk}*INST_k \\
& + \gamma_{122}*HSUP_{ijk}*STUD_{jk}*COMP_k + \gamma_{130}*HSUP_{ijk}*NUMB_{jk} \\
& + \gamma_{131}*HSUP_{ijk}*NUMB_{jk}*INST_k + \gamma_{132}*HSUP_{ijk}*NUMB_{jk}*COMP_k \\
& + \gamma_{140}*HSUP_{ijk}*ALGB_{jk} + \gamma_{141}*HSUP_{ijk}*ALGB_{jk}*INST_k \\
& + \gamma_{142}*HSUP_{ijk}*ALGB_{jk}*COMP_k + \gamma_{150}*HSUP_{ijk}*PREP_{jk} \\
& + \gamma_{151}*HSUP_{ijk}*PREP_{jk}*INST_k + \gamma_{152}*HSUP_{ijk}*PREP_{jk}*COMP_k \\
& + \gamma_{160}*HSUP_{ijk}*HMWC_{jk} + \gamma_{161}*HSUP_{ijk}*HMWC_{jk}*INST_k \\
& + \gamma_{162}*HSUP_{ijk}*HMWC_{jk}*COMP_k + \gamma_{170}*HSUP_{ijk}*LOAD_{jk} \\
& + \gamma_{171}*HSUP_{ijk}*LOAD_{jk}*INST_k + \gamma_{172}*HSUP_{ijk}*LOAD_{jk}*COMP_k \\
& + \gamma_{180}*HSUP_{ijk}*DEGR_{jk} + \gamma_{181}*HSUP_{ijk}*DEGR_{jk}*INST_k \\
& + \gamma_{182}*HSUP_{ijk}*DEGR_{jk}*COMP_k + r_{0jk} + r_{1jk}*HSUP_{ijk} + u_{00k} + u_{01k}*HWFR_{jk}
\end{aligned}$$

$$+ u_{02k} * STUD_{jk} + u_{03k} * NUMB_{jk} + u_{04k} * ALGB_{jk} + u_{05k} * PREP_{jk}$$

$$+ u_{06k} * HMWC_{jk} + u_{07k} * LOAD_{jk} + u_{08k} * DEGR_{jk} + u_{10k} * HSUP_{ijk}$$

$$+ u_{11k} * HSUP_{ijk} * HWFR_{jk} + u_{12k} * HSUP_{ijk} * STUD_{jk} + u_{13k} * HSUP_{ijk} * NUMB_{jk}$$

$$+ u_{14k} * HSUP_{ijk} * ALGB_{jk} + u_{15k} * HSUP_{ijk} * PREP_{jk} + u_{16k} * HSUP_{ijk} * HMWC_{jk}$$

$$+ u_{17k} * HSUP_{ijk} * LOAD_{jk} + u_{18k} * HSUP_{ijk} * DEGR_{jk} + e_{ijk} \quad (3.26)$$

3.6.4 Centering

Centering is mostly preferred in order to estimate the intercept more precisely for HLM and comment about output more meaningfully. Especially, when there is not an accurate zero for independent variables, centering is helpful to eliminate the multicollinearity problem between independent variables (Newman, Newman & Salzman, 2010). Centering is done mostly to improve the parameter estimation of analysis (Kreft, Leeuw & Aiken, 1995).

For HLM analysis, raw scores might be used or centering might be done for group means or grand means. However, whether the centering is needed or not, or if centering is needed which type of centering is more appropriate might differ according to the specified model or purposes of the analysis (Paccagnella, 2006; Newman, Newman & Salzman, 2010). Generally, for nested data structures group mean centering is suggested for subset group, whereas the grand mean centering is suggested for higher set. For the circumstances aiming to distinguish group-level effects from individual characteristics, it should be preferred to center the variables according to group means is desirable (Paccagnella, 2006). When in fact, there is not an accurate consensus about centering the data for HLM (Kreft, Leeuw & Aiken, 1995).

In this research, parallel with the aim of the research, at student level group mean centering and for school and country levels grand mean centering had been used.

3.6.5 Parameter estimation: Maximum likelihood

As mentioned at Raudenbush, Bryk, Cheong, Congdon, & du Toit (2004), there are three kinds of parameter estimation are applicable, namely "empirical Bayes estimates of randomly varying level-1 coefficients, generalized least squares estimates of the level-2 coefficients and maximum-likelihood estimates of the variance and covariance components" (p.9). During the current study, fully maximum likelihood is used for parameter estimation.

3.6.6 Sample size

The proper sample size is very important for HLM to have parameter estimates eliminated from random error as much as possible and to get acceptable analysis power. In an ideal circumstance, it is desirable to have as many sample size as possible at each level of the analysis (Snijders, 2005). For two-level HLM, the sample size is questioned for both levels. According to Maas and Hox (2005), the second-level should have less than 100 units to get small amount of error. On the other hand, Bassiri (1988) stated that for two-level hierarchical linear models 30 groups with 30 individual at each (so 900 in total) is need for a fair sample group (as cited by Hofmann, 1997). On the other hand, for three-level HLM, to the extent possible search, there is not a concurrent sample size on the literature. Therefore, at least 30 cases at the highest level and each case to have at least 30 lower level cases requirement is followed for the current study. Therefore, after all the preliminary

analysis would be done, 36 countries, 6,586 schools and 219,788 students are considered as enough sample size for three-level HLM analysis.

First of all, the missing values and extreme values will be eliminated from the data set before the assumption tests for HLM would have been done.

3.6.7 Missing data

HLM software programs are able to deal with missing data on level-1 file. The programs suggest ways to deal with missing data, namely:

1. Deleting the cases according to list at MDM file is formed
2. Deleting the cases according to list when the analysis is run
3. Analysis of multiply-imputed data (Raudenbush et al., 2004, p.46; Raudenbush et al., 2011, pp. 43-44)

During the current study, level-1 missing data were dealt as deleting list wise during the MDM file was formed. On the other hand, at level-2 and level-3 HLM expect a complete data set. It is suggested to clean the level-2 and level-3 data from missing cells before starting to the analysis; otherwise HLM programs delete level-2 units automatically list wise. This can cause many data to be lost (Raudenbush et al., 2004). For this reason, level-2 and level-3 variables had been checked for missing values before starting to the analyses. TIMSS 2015 eighth grade administration has 39 countries attended, whereas Egypt, Russian Federation and Israel were deleted from data set and analyses are completed for 36 countries, 6,586 schools and 219,788 students. Egypt was deleted because level-3 variables were completely missing. Russian Federation is deleted because five out of 11 predictor variables at level-2 were completely missing. At last, Israel is deleted by HLM program because the predictor variable home study support (HSUP) was completely missing. Before the

analysis, missing data analysis was done again and level-3 did not have missing data. At Level-2, there were some missing values for some variables, but all were less than 5% of the data (Tabachnick & Fidell, 2001). Because of the limitation of HLM software that it cannot handle with missing data at level-2, those missing data are replaced with series mean.

3.6.8 Extreme values

Extreme values had been examined for only Level-1 variables. For one way extreme values, the z normal points were examined for both level variables. At the end, 226 students were deleted from the data set because of extreme values which exceeds (-3, +3) interval on z standardized values.

3.6.9 The assumptions of hierarchical linear modeling

The lower level units (classroom, teacher, student etc.) are nested within upper level units (country, school, city etc.). First of all, as a methodological assumption which needs to be considered is that the explanatory variables might belong to more than one level whereas the dependent variable can only be belong to the first which is the lower level (Hofmann, Griffin & Gavin, 2000).

Due to the nature of the data set, in order to get fair results HLM require statistical assumptions. For three-level HLM, following statistical assumptions need to be provided (Subedi, 2005; Raudenbush et. al. 2011):

1. Outcome variable has a nonlinear relation with level-1 predictor variables, whereas linear relations with level-2 and level-3 predictor variables.

2. The model is correctly formulated, which means important variables are included, unrelated variables are excluded in the model and observed variables are measured without an error.
3. The residuals at both level-2 and level-3 are independently and multivariate normally distributed.
4. The predictor variables do not have multicollinearity with each other.

These assumptions were checked before starting to the real HLM analysis (see Appendix A, B, C). For examining the model fit and validity of assumptions, HLM3 produces three residual files at each level of level 1, level 2 and level 3. Basically, the level-1 residual file contains the differences between the observed and fitted values of each level-1 units, the square root of σ^2 and observed values of each level-1 variable. For level-1 residuals both P-P and Q-Q plots have been checked (see Appendix B), and the plot appeared as linear, which suggest there is not a serious departure from a normal distribution and so the assumption is justifiable.

Level-2 residual file includes level-2 and level-3 ID numbers, followed by number of level-1 units nested in level-2 unit j in the level-3 unit k. Then two empirical Bayes (EB) residuals (one for the intercept and one for HSUP), two ordinary least squares (OLS) residuals for the same quantities, the fitted values and EB coefficients followed by the posterior variances and covariance of the EB estimates (pv2_0_0, pv2_1_0, pv2_1_1). Those can be seen on Appendix B and C that both the descriptive statistics of Level-2 residual file (Table C1) and the examination of EB versus OL residuals for level-1 predictor variable (Figure B4) are calculated. Similar as expected on Raudenbush et al. (2011, p. 41) the EB residuals for the slope of HSUP are much more compact than the OL residuals. While the OL residuals range between -100,000 to +100,000, the range for the EB residuals is -

10,000 to +10,000. It is noted that HLM3 does not request the test of level-1 homogeneity. Similar variables are observed in level-3 residual file (Table C2-C3). Residual files had been examined according to the given guide on Raudenbush et al. (2011) and considered as the assumptions are tenable for the HLM3.

CHAPTER 4

RESULTS

In this chapter of the thesis, firstly descriptive statistics of the data and then findings are presented in the order of model development of three-level HLM.

4.1 Basic descriptive statistics

The final data which has been based on TIMSS 2015 eighth grade administration data contained 219,788 students nested within 6,586 schools in 36 countries. As a representative of student socio-economic status, number of home study support index which is derived by IEA is used as the predictive variable. As dependent variable, students' mathematics achievement which is the 1st plausible value of mathematics scores at 8th grade is considered. Detailed descriptive statistics about student level variables are presented in Table 8.

Table 8. Student-Level Variables (N=219,788)

Variable	%	Mean	SD	Minimum	Maximum
Predictor Variable:					
Number of home study support	…	1.52	0.64	0	2
0=Neither own room nor internet connection	8.3	…	…	…	…
1=Either own room or internet connection	32.1	…	…	…	…
2=Both own room and internet connection	59.7	…	…	…	…
Dependent Variable:					
Mathematics achievement		474.85	106.18	150.57	803.47

The predictor variable, namely number of home study support, is derived from the answers to the question "Do you have any of these things in your home?" which has answers recoded as 0= "Neither own room now internet connection",

1= “Either own room or internet connection”, 2= “Both own room and internet connection”. As seen on Table 8, the variable number of home study support has minimum of 0, maximum of 2, mean of 1.52 and standard deviation (SD) of 0.64. From the student sample of the study (N=219,788), only 8.3% has neither own room nor internet connection whereas 59.7% has both own room and internet connection. For dependent variable mathematics achievement, minimum score is 150.57 and maximum score is 803.47. Mean of whole sample is 474.85 with SD of 106.18. It is seen that the average calculated from the students’ mathematics scores of all countries’ is below TIMSS Scale Center Point 500 as being 474.85.

The summary descriptive statistics for school level (level-2) predictors are presented in Table 9.

Table 9. School-Level Predictor Variables (N=6,586)

Variable	%	Mean	SD	Minimum	Maximum
Teachers’ BS degree	…	2.54	1.09	0.00	4.00
0= No formal education beyond upper secondary	2.7	…	…	…	…
1=All other majors	11.4	…	…	…	…
2=Major in math but no major in math education	41.6	…	…	…	…
3= Major in math education but no major in math	17.0	…	…	…	…
4= Major in math or math education	27.3	…	…	…	…
Homework frequency		3.76	1.03	1.00	5.00
1=No mathematics homework	2.7	…	…	…	…
2=Less than once a week	6.8	…	…	…	…
3=One or 2 times a week	29.7	…	…	…	…
4=Three or 4 times a week	32.4	…	…	…	…
5= Every day	28.3	…	…	…	…
Number of students in the classes	…	27.49	9.99	1.00	95.00
Weekly mathematics class time	…	224.43	65.36	3.00	600.00
The topic ‘Numbers’ taught to TIMSS classes	…	0	1	-7.87	.65
The topic ‘Algebra’ taught to TIMSS classes	…	0	1	-2.93	1.28
The topic ‘Geometry’ taught to TIMSS classes	…	0	1	-3.86	1.12
The topic ‘Data and Chance’ taught to TIMSS classes	…	0	1	-1.61	1.07
Preparation to teach mathematics	…	0	1	-3.43	1.08
Mathematics homework for the TIMSS class	…	0	1	-4.01	1.92
Working load of teacher	…	0	1	-3.01	2.91

Teacher' BS degree is a derived variable by IEA and recoded for this study according to the research purposes. As seen on Table 9, the largest percentage of teachers' BS degree (41.6%) belongs to the category of "Major in math but no major in math education" whereas "Major in math education but no major in math" category has only 17% of whole sample. Homework frequency is a five point Likert item, which has minimum point 1 (No mathematics homework) and maximum point 5 (Every day). Homework frequency has mean 3.76 with SD 1.03, which shows also with percentages that most of the teachers in schools (60.7%) assign homework at least three times a week. Number of students in the classes and weekly mathematics class time are continuous variables which were used as the way they were at TIMSS 2015 data set. Weekly mathematics class time is a derived variable which is formed by IEA. Following four variables on the Table 9 which are about the topics taught to TIMSS classes are derived variables by IEA, and the remaining three variables had been derived during the current research and taken as z standard scores. Therefore, these seven variables have 0 as mean and 1 as SD.

The summary statistics for country level predictor variables are presented in Table 10.

Table 10. Country-Level Predictor Variables (N=36)

Variable	Mean	SD	Minimum	Maximum
Instructional Time	137.44	23.48	99.00	194.00
Number of TIMSS mathematics topics intended to be taught	17.37	3.36	3.00	20.00
Mathematics and numeracy skills at pre-primary education	.53	.51	0	1
Classroom tracking/grouping by achievement	.28	.45	0	1
Components prescribed by mathematics curriculum	.06	.99	-1.45	.84
Methods used to evaluate the implementation of mathematics	.03	1.01	-2.17	1.17

Country instructional time decided for mathematics classes and number of TIMSS mathematics topics intended to be taught at the end of eighth grade are continuous variables, have means and SDs as 137.44 (23.48) and 17.37 (3.36), minimum values 99 and 3, maximum values 194 and 20 respectively. The dichotomous predictor variables, Mathematics and numeracy skills at pre-primary education and Classroom tracking/grouping by achievement have mean 0.53 and 0.28, and SDs 0.51 and 0.45 respectively. Components prescribed by mathematics curriculum and Methods used to evaluate the implementation of mathematics are variables which had been derived during the current research and used as z standard scores.

Development of the Three-Level Model

At the beginning of the research, one predictor variable for the student level model, 11 for the school level and six for the country level are decided. During the development of the model, firstly, predictor variables are tested as conditional for their own level and unconditional for other levels. In other words, level-1 predictor variable was added to HLM3 solely and tested level-2 predictor variables were added to model without other level predictor variables and tested. Eventually, level-3 predictor variables were added to the model solely and tested. The purpose of these processes is to conceive the proportions of variance at each level. Furthermore, the model which takes only level-1 predictor variable as conditional and takes level-2 and level-3 variables as unconditional forms the model 2 after the unconditional model which is model 1 in the development of the three-level hierarchical linear model. Model 3 is formed as adding the level-2 predictors to model 2 and final

model is formed as adding the level-3 predictors to model 3. Hence, development of the three-level hierarchical model is completed.

In order to select the potential predictor variables, exploratory analysis was used in the development of the model. As considered the t-statistic on level-2 and level-3 exploratory analysis results, the variables which have absolute t-statistics larger than 2.0 (Subedi, 2005) were considered as possible predictor variables.

4.2 Results for unconditional model

In order to answer to the first research question, the unconditional model of hierarchical linear modeling (HLM) is performed. In this model, there is not any predictor variable stated at any of three-levels. The purpose of this model is to reveal how the total variance is distributed between three levels (student, school and country) of HLM (Raudenbush, Bryk, & Congdon, 2011).

The estimated fixed and random effects based on HLM3 output are given in Table 11.

Table 11. HLM3 Analysis Results of Model 1 (Not Any Predictor Included)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
School model for student level intercept, π_0					
Country model for school level intercept, β_{00}					
Intercept, γ_{000}	483.19	10.82	44.66	35	<0.001**
Random Effect	**Standard Deviation**	**Variance Component**	***d.f.***	**χ^2**	***p*-value**
Schools (level-2), r_0	51.5333	2655.6789	6550	123644.66583	<0.001**
Students (level-1), e	69.1997	4788.6053			
Countries (level-3), u_{00}	64.7742	4195.6950	35	10107.31461	<0.001**

** Statistically significant at p <.01

As seen on Table 11, the intercept coefficient (483.19) is interpreted as the average mathematics achievement score. The symbols, π_0, β_{00} and γ_{000} refer school, country and general means respectively. The general mean (γ_{000}) is estimated as 483.19 with 10.82 standard error (SE) approximately. It is expected that the real value of the general mean to be between 461.98 and 504.40 (confidence interval of 95% γ_{000}=483.19±(1.96)(10.82)). The intercept coefficient (483.19) is found as significant ($p<0.001$), which means that mathematics achievement scores varied significantly at all three levels (student, school and country).

Additionally, it is expected that 95% of schools' average mathematics achievement scores to be between 382.19 and 584.19 ($\widehat{\gamma_{000}} = 483.19 \pm (1.96)(51.53)$. Similarly, it is expected that 95% of countries' average mathematics achievement scores to be between 356.24 and 610.14 ($\widehat{\gamma_{000}} = 483.19 \pm (1.96)(64.77)$. Both the level-2 and level-3 variance components are observed as significant ($p<0.001$). When the random effects are examined, the variability of mathematics achievement scores are estimated at within school as (e) 4788.60, between schools (r_0) as 2655.68 and between countries (u_{00}) as 4195.69. These significant values show that differences between average mathematics achievement scores are statistically significant at both school and country level.

In order to calculate, the amounts of variations at each level, the equations of proportion of variances over three levels are as follows (http://www.ssicentral.com/hlm/help6/faq/Proportion_Variance_Explained__Three-Level_Models.pdf).

σ^2= var(e_{ijk}), represents the variability of level-1 units (students) in outcome

τ_π= var(r_{0jk}), represents the variability of level-2 units (schools) in outcome

τ_β= var(u_{00k}), represents the variability of level-3 units (countries) in outcome

The proportion of variance over level-1 units=$\frac{\sigma^2}{\sigma^2+\tau_\pi+\tau_\beta}$ (4.1)

The proportion of variance over level-2 units=$\frac{\tau_\pi}{\sigma^2+\tau_\pi+\tau_\beta}$ (4.2)

The proportion of variance over level-3 units=$\frac{\tau_\beta}{\sigma^2+\tau_\pi+\tau_\beta}$ (4.3)

Hence, the proportions of variances for the current study are calculated as below:

The proportion of variance within school:

$$\hat{p}=\frac{4788.6053}{4788.6053+2655.6789+4195.6950}=0.411$$

The proportion of variance between schools:

$$\hat{p}=\frac{2655.6789}{4788.6053+2655.6789+4195.6950}=0.228$$

The proportion of variance between countries:

$$\hat{p}=\frac{4195.6950}{4788.6053+2655.6789+4195.6950}=0.360$$

According to these results, 41.1% of the variance in TIMSS mathematics achievement scores is derived from student level, 22.8% of the variance is derived from school level and 36.0% of the variance is derived from country level. The amount of the proportion of variance is highest at student level, which means that the differences between mathematics achievement scores are based on student level variables at most.

On the other hand, included country level variables explain higher amount of variance than school level predictor variables. Since mathematics achievement scores vary significantly at each of three levels, the consequent models are estimated using the exploratory analysis. For the findings of the following research question, firstly only school level predictor variables and then only the country level predictor

variables are added to the model according to the results of exploratory analysis of country level predictor variables.

The reliability estimates of level-2 and level-3 predicting outcome models are calculated as 0.936 and 0.996 respectively, which are accepted as very high reliability coefficients. Raudenbush et al. (2011) mentioned that very low reliabilities, for instance <0.10, point out a random coefficient might be regarded as fixed in following analyses. This has not been the case during all the analyses including the final model.

4.3 Results for only school level predictors included model (Model 1a)

In this step, based on the significant differences between students' average mathematics achievements on country levels, as taking level-1 and level-3 unconditional (controlled), the level-2 conditional model is formulated. In this model, level-1 and level-3 regression coefficients (π_{0jk} and γ_{000}) are assumed as randomly varying and there is not any level-1 or level-3 slope coefficient. At level-2, possible level-2 predictors are added to the model in order to explain the differences between country means. Before starting to the analysis exploratory analysis is performed with eleven predictors, namely teachers' BS degree, homework frequency, number of students in the classes, weekly mathematics class time, the topics 'Numbers' taught to TIMSS classes, the topics 'Algebra' taught to TIMSS classes, the topics 'Geometry' taught to TIMSS classes, the topics 'Data and Chance' taught to TIMSS classes, preparation to teach mathematics, mathematics homework for the TIMSS class and working load of teachers. According to the t-statistics criteria of $|t| \geq 2.0$, eighth out of eleven variables are included as level-2 predictor variables. These variables are homework frequency (t=6.511), number of students in

the classes (t=1.918), the topics 'Numbers' taught to TIMSS classes (t=3.780), the topics 'Algebra' taught to TIMSS classes (t=3.301), preparation to teach mathematics (t=2.569), mathematics homework for the TIMSS class (t=2.063), working load of teachers (t=-7.137) and teachers' BS degree (t=4.017). The variable number of students in the classes has a t-value which is below but close to the value 2.0, in order to investigate the interaction effects at further steps of model development, that variable is also taken into the models.

In this step, considering mathematics achievement (MATH) as the outcome variable, the model which includes only possible level-2 predictor variables is formed. The estimated fixed and random values, based on HLM3 output are given in Table 12.

Table 12. HLM3 Analysis Results of Only Level-2 Predictor Variables Included

Fixed Effect	Coefficient	SE	t-ratio	p-value	
Intercept, γ_{000}	483.07	10.81	44.70	<.001**	
Level 2 Predictors:					
Homework frequency, γ_{010}	4.99	1.57	3.18	.003**	
Number of students in the classes, γ_{020}	0.94	0.29	3.30	.002**	
The topics 'Numbers' taught to TIMSS classes, γ_{030}	1.83	1.18	1.55	.131	
The topics 'Algebra' taught to TIMSS classes, γ_{040}	3.51	1.55	2.27	.030*	
Preparation to teach mathematics, γ_{050}	1.20	1.17	1.03	.311	
Mathematics homework for the TIMSS class, γ_{060}	1.32	0.84	1.56	.126	
Working load of teachers, γ_{070}	-4.58	1.07	-4.28	<.001**	
Teachers'BS degree, γ_{080}	1.39	0.85	1.65	.109	
Random Effect	SD	Variance Compo-nent	df	X^2	p-value
Schools (level-2), r_0	48.08161	2311.842	6105	100803.939	<0.001**
Students (level-1), e	69.19902	4788.504			
Countries (level-3), u_{00}	64.72128	4188.844	34	10311.3258	<0.001**

*Statistically significant at p<.05.
** Statistically significant at p <.01
***p-value with robust standard errors is p<.05.

At Table 12, level-3 interception (γ_{000}) is found approximately 483.07 with 10.81 standard error (SE). It is expected that the real value of the general mean to be between 472.26 and 493.88 (95% γ_{000}=483.07±(1.96)(10.81)). The intercept coefficient (483.07) is significant ($p<0.001$).

When the significance of predictor variables are examined, it is seen that the variables "homework frequency" (p=.003), "number of students in the classes" (p=.002) and "working load of teacher" (p<.001) are appeared as significant (p<.01), and "the topic 'Algebra' taught to TIMSS classes" (p=.030) is significant at p<.05 level. The coefficient of "homework frequency" (γ_{010}) is estimated as 4.99 with standard error 1.57. That coefficient shows that when a school's "homework frequency" increases one standard point, school's average mathematics achievement point will increase about 5 points. In other words, homework frequency is positively related with mathematics achievement scores. Additionally, "number of students in the classes" shows also a positive relation with mathematics achievement scores. The coefficient of "number of students in the classes" (γ_{020}) is estimated as 0.95 with standard error 0.29. That coefficient shows that when a school's "number of students in the classes" increases one standard point, school's average mathematics achievement point will increase about 1 point. A similar result is found also for "the topic 'Algebra' taught to TIMSS classes". The coefficient of "the topic 'Algebra' taught to TIMSS classes" (γ_{040}) is estimated as 3.51 with standard error 1.55. That coefficient shows that when a school's "the topic 'Algebra' taught to TIMSS classes" index increases one standard point, school's average mathematics achievement point will increase about 3.5 point. In other words, a school's mathematics average score is positively related with the percentage of students taught Algebra.

On the other hand, the coefficient of "working load of teachers" (γ_{070}) is estimated as -4.58 with standard error 1.07. That coefficient shows that when a school's "working load of teachers" index increases one standard point, school's average mathematics achievement point will decrease about 4.5 point. In other words, a school's mathematics average score is negatively related with the working load of teachers.

When the random effects are examined, the variability of mathematics achievement scores are estimated at within school as (e) 4788.50, the variability between schools after the mentioned eighth variables are controlled (r_0) as 2311.84 and the variability of between countries (u_{00}) is estimated as 4188.84. These significant values show that differences between average mathematics achievement scores are statistically significant at both school and country level after the school level predictor variables are controlled.

The explained variance at level-2 can be calculated as comparing the estimated variance at unconditional model and the model with only level-2 predictor variables are conditional.

$$\text{The explained variance at level-2} = \frac{\tau_\pi(\text{Model 1}) - \tau_\pi(\text{Model 1a})}{\tau_\pi(\text{Model 1})}$$

$$= \frac{2655.68 - 2311.84}{2655.68} = 0.13$$

According to the result of explained variance at level-2, approximately 13% of the variance of between countries mathematics achievement scores is explained by the predictor variables homework frequency, number of students in the classes, the topics 'Numbers' taught to TIMSS classes, the topics 'Algebra' taught to TIMSS classes, preparation to teach mathematics, mathematics homework for the TIMSS class, working load of teachers and teachers' BS degree. It is obtained at previous

model that school level explains 23% of mathematics achievement scores, so it can be stated that approximately 3% (0.23*0.13=0.03) of the total variance at mathematics achievement is sourced from the mentioned eighth school level predictor variables. The following estimated proportion refers the remaining unexplained variance of mathematics achievement scores after the mentioned predictor variables are controlled.

$$\text{The proportion of variance over level-2 units}=\frac{\tau_\pi}{\sigma^2+\tau_\pi+\tau_\beta}$$

$$=2311.84/(4788.50+2311.84+4188.84)$$

$$=0.205$$

Hence, at last status, after the eighth level-2 predictor variables are controlled, approximately 20.5% of the variance between countries remains as unexplained by those variables. The reliability estimates of level-2 and level-3 outcome models are calculated as 0.928 and 0.996 respectively, which are very high reliability coefficients.

4.4 Results for only country level predictors included model (Model 1b)

In this step, based on the significant differences between students' average mathematics achievements on country levels, as taking level-1 and level-2 unconditional (controlled) the level-3 conditional model is formulated. In this model, level-1 and level-2 regression coefficients (π_{0jk} and β_{00k}) are assumed as randomly varying and there is not any level-1 or level-2 slope coefficient. At level-3, possible level-3 predictors are added to the model in order to explain the differences between country means. Before starting to the analysis exploratory analysis is performed with six predictors, namely components prescribed by mathematics curriculum, instructional time, number of TIMSS mathematics topics intended, methods used to

evaluate the implementation of mathematics curriculum, mathematics and numeracy skills at preprimary education, classroom tracking/grouping by achievement. According to the t-statistics criteria of $|t| \geq 2.0$, two out of six variables are included as level-3 predictor variables. These variables are instructional time (t=-2.239) and components prescribed by mathematics curriculum (t=-1.676). The later variable has a t-value which is below but close to the value 2.0, in order to investigate the interaction effects at further steps of model development, the variable components prescribed by mathematics curriculum is also taken into the models.

In this step, considering mathematics achievement (MATH) as the outcome variable, the model which includes only the possible level-3 predictor variables is formed. The estimated fixed and random values, based on HLM3 output are given in Table 13.

Table 13. HLM3 Analysis Results of Only Level-3 Predictor Variables Included

Fixed Effect	Coefficient	SE	t-ratio	p-value	
Intercept, γ_{000}	483.19	9.52	50.73	<.001**	
Level 3 Predictors:					
Model for country mean (β_{00}):					
Components prescribed by mathematics curriculum***γ_{002}	-18.52	9.76	-1.90	.067 (.037*)	
Instructional time γ_{001}	-1.06	0.41	-2.57	.015*	
Random Effect	SD	Variance Compo-nent	df	X^2	p-value
Schools (level-2), r_0	51.53453	2655.8081	6550	123645.595	<0.001**
Students (level-1), e	69.19972	4788.6017			
Countries (level-3), u_{00}	56.99414	3248.3316	33	7565.490	<0.001**

*Statistically significant at p<.05.
** Statistically significant at p <.01
***p-value with robust standard errors is p<.05.

At Table 13, level-3 interception (γ_{000}) is found approximately 483.19 with 9.52 standard error (SE). It is expected that the real value of the general mean to be

between 464.53 and 501.85 (95% γ_{000}=483.19±(1.96)(9.52)). The intercept coefficient (483.19) is significant (p<0.001). The coefficient of "components prescribed by mathematics curriculum" (γ_{002}) is estimated as -18.52 with standard error 9.76. That coefficient shows that when country's "components prescribed by mathematics curriculum" index increases one standard point, country's average mathematics achievement point will decrease about 18 points. In other words, in the current research context, because higher score refers more centralized education, more centralized education systems effect to mathematics achievement score average negatively. The effect of countries' "components prescribed by mathematics curriculum" index on the average mathematics achievement scores is statistically significant at p<.05 (p=.035) level with robust standard error, which means that this curriculum variable has different impact on the students and schools at different countries.

The coefficient of "instructional time" (γ_{001}) is estimated as -1.06 with standard error 0.41. That coefficient shows that when country's amount of "instructional time" increases one point (1 hour) country's average mathematics achievement point will decrease about 1 point. In other words, in the current research context, because higher score refers more time spent on mathematics lessons, spending more time on mathematics lessons effects mathematics achievement score average negatively at a small amount. The effect of countries' "instructional time" on the average mathematics achievement scores is statistically significant at p<.05 level, which means that this curriculum variable has different impact on the students and schools at different countries.

When the random effects are examined, the variability of mathematics achievement scores are estimated at within school as (e) 4788.60, the variability

between schools (r_0) as 2655.81 and the variability of between countries after the predictor variables "components prescribed by mathematics curriculum" and "instructional time" are controlled (u_{00}) is estimated as 3248.33. These significant values show that differences between average mathematics achievement scores are statistically significant at both school and country level after the country level predictor variables are controlled.

Additionally, it is expected that 95% of schools' average mathematics achievement scores to be between 382.19 and 584.19 ($\widehat{\gamma_{000}} = 483.19 \pm (1.96)(51.53)$ after the predictor variables at level-3 are controlled. Because only the level-3 predictor variables are controlled, school level variance is changed slightly (less than a fraction of percent). Similarly, it is expected that 95% of countries' average mathematics achievement scores to be between 371.49 and 594.89 ($\widehat{\gamma_{000}} = 483.19 \pm (1.96)(56.99)$. This interval is narrower than the interval which is calculated for the model without controlling the level-3 predictors, which means that controlling predictor variables causes more accurate estimations about average achievements.

In order to examine, whether the variance of countries' average mathematics achievement is different than zero after the level-3 predictor variables are controlled, the p value of u_{00} is observed. Because the p value is smaller than 0.05 ($p<0.001$), it can be interpreted that after the level-3 predictor variables are controlled, significant variance is remained without being explained in countries' average mathematics achievement scores. The explained variance at level-3 can be calculated as comparing the estimated variance at unconditional model and the model with only level-3 predictor variables are conditional.

$$\text{The explained variance at level-3}=\frac{\tau_\beta(\text{Model 1})-\tau_\beta(\text{Model 1b})}{\tau_\beta(\text{Model 1})}$$

$$=\frac{4195.69-3248.33}{4195.69}= 0.23$$

According to the result of explained variance at level-3, approximately 23% of the variance of between countries mathematics achievement scores is explained by the predictor variables “components prescribed by mathematics curriculum” and “instructional time”. It is obtained at previous model that country level explains 36% of mathematics achievement scores, so it can be stated that approximately 8% (0.36*0.23=0.08) of the variance at mathematics achievement is sourced from the predictor variables “components prescribed by mathematics curriculum” and “instructional time”. The following estimated proportion refers the remaining unexplained variance of mathematics achievement scores after the mentioned predictor variables are controlled.

The proportion of variance over level-3 units$=\frac{\tau_\beta}{\sigma^2+\tau_\pi+\tau_\beta}$

=3248.33/(2655.81+4788.60+3248.33)

=0.303

Hence, at last status, after the predictor variables “components prescribed by mathematics curriculum” and “instructional time” are controlled, approximately 30.3% of the variance between countries remains as unexplained by those variables. The reliability estimates of level-2 and level-3 outcome models are calculated as 0.936 and 0.995 respectively, which are very high reliability coefficients.

4.5 Results for only student level predictor included model (Model 2)

In this step, based on the significant differences between students’ average mathematics achievements on student, school and country levels, the level-1 conditional, level-2 and level-3 unconditional models are formulated. It is expected

that as one of the indicators of student socio-economic status, home study support would be an important student level predictor variable. Based on the meaningful correlation between the dependent variable mathematics achievement and the predictor variable home study support (r=.257, p<.001), home study support is included into the further analysis for the development of three-level model as student level predictor variable. The school and country variance components are included or excluded in following model if the variance components are significant at p<.05 (Subedi, 2005).

The final estimation of fixed effects and random effects for model 2 which includes only student level predictor variable as conditional are presented in Table 14, based on HLM3.

Table 14. HLM3 Analysis Results of Model-2 (Only Level-1 Predictor Variables Included)

Fixed Effect	Coefficient	SE	t-ratio	d.f.	p-value
School model for student level intercept, π_{0jk}					
Country model for school level intercept, β_{00k}					
Intercept, γ_{000}	483.18	10.82	44.66	35	<.001**
School model for student level effect of home study support, π_{1jk}					
Country model for school level intercept, β_{10k}					
Intercept, γ_{100}	3.91	1.18	3.32	35	.002**
Random Effect	SD	Variance Component	d.f.	χ^2	p-value
Schools (level-2), r_0	51.5479	2657.1895	6364	123098.1048	<0.001**
Home study support slope, r_1	8.2829	68.6066	6364	7177.6203	<0.001**
Students (level-1), e	68.9431	4753.1455			
Countries (level-3), u_{00}	64.7767	4196.0234	35	10100.4111	<0.001**
Home study support/Interception2, u_{10}	6.7253	45.2298	35	461.8279	<0.001**

** Statistically significant at p <.01

At Table 14, the interception (γ_{000}) is found as approximately 483.18 with 10.82 standard error (SE). The intercept coefficient (483.18) is significant (p<0.001).

The coefficient of "home study support" (γ_{100}) is estimated as 3.91 with standard error 1.18. That coefficient shows that when students' "home study support" value increases one point students' average mathematics achievement point will also increase about 4 points. In other words, home study support is positively related with mathematics achievement score average.

When the random effects are examined, after the predictor variable "home study support" is controlled, the variances of mathematics achievement scores are significant. These significant values show that differences between average mathematics achievement scores are statistically significant at both student, school and country levels after the student level predictor variable is controlled ($p<0.001$). It can be interpreted that after the level-1 predictor variable, home study support, is controlled, significant variance is remained without being explained in countries' average mathematics achievement scores.

The explained variance at level-1 can be calculated as comparing the estimated variance at unconditional model and the model with only level-1 predictor variables are conditional.

$$\text{The explained variance at level-1}=\frac{\sigma^2(\text{Model 1})-\sigma^2(\text{Model 2})}{\sigma^2(\text{Model 1})}$$

$$=\frac{4788.60-4753.14}{4788.60}=0.007$$

According to the result of explained variance at level-1, approximately 0.7% of the variance of between countries mathematics achievement scores is explained by the predictor variable "home study support". It is obtained at previous model that student level explains 41% of mathematics achievement scores, so it can be stated that approximately 0.3% (0.41*0.007=0.003) of the variance at mathematics achievement is sourced from the predictor variable "home study support". The

following estimated proportion refers the remaining unexplained variance of mathematics achievement scores after the predictor variable “home study support” is controlled.

The proportion of variance over level-1 units=$\frac{\sigma^2}{\sigma^2+\tau_\pi+\tau_\beta}$

=4753.15/(4753.15+2657.19+4196.02)

=0.409

Hence, at last status, after the predictor variable “home study support” is controlled, approximately 40.9% of the variance between countries remains as unexplained by that variable. The estimations of reliabilities for estimating student and school level model outcomes are given in Table 15.

Table 15. Final Estimation of Level-1 and Level-2 Reliabilities

Coefficient	Reliability estimate
Level 1 Intercept, π_0	0.938
Home study support, π_1	0.121
Level 2 Intercept, β_{00}	0.996
Home study support/ Intercept, β_{10}	0.904

In sum, the findings of significant ($p<.01$) intercepts in student level suggests the development of the model as specifying school-level variables as extending the model 2. In order to consider the level-2 predictor variables, first an exploratory analysis was done and according to the exploratory analysis, seven predictor variables (HWFR, NUMB, ALGB, PREP, HMWC, LOAD and DEGR) had t-values above 2.0. Additionally, the model which considers only the level-2 predictors as conditional was applied and three variables (HWFR, STUD and LOAD) had significant fixed effects. The exploratory analysis outputs are given at Appendix D. Therefore, for further model analysis, eighth level-2 predictor variables (HWFR,

STUD, NUMB, ALGB, PREP, HMWC, LOAD and DEGR) and one student level predictor variable (HSUP) would be tested.

4.6 Results for student and school level predictors included model (Model 3)

In this step of the research, student and school levels are taken as conditional and country level is as unconditional and formulated in order to examine how the effect of student level predictor variable depend on the school level predictor variables. As mentioned at previous section, one student-level predictor variable (home study support) and eighth school-level predictor variables (homework frequency, number of students in the classes, the topic 'numbers' taught to TIMSS classes, the topic 'algebra' taught to TIMSS classes, teachers' preparedness to teach math, math homework usage index, working load of teachers, teachers' BS degree) are considered while performing analysis for model 3. All school level variance components are included in the model based on their significant values stated at previous step.

The final estimation of fixed effects for model 3 which includes student level and school level predictor variables as conditional and country level predictor variables as unconditional are presented in Table 16, based on HLM3.

As seen on Table 16, model 3, the conditional model for Level-1 and level-2 variables, is more complex than model 2 which has only Level-1 variables as controlled. Firstly, mathematics achievement scores are directly related to the student characteristic, home study support. Secondly, effects of this variable are moderated by school variables, namely number of student in the classes and teachers' BS degree. In addition, three out of eighth school characteristics are directly related to mathematics achievement scores.

Table 16. HLM3 Analysis Results of Model 3-Fixed Effects (Level-1 and Level-2 Variables Included)

Fixed Effect	Coefficient	SE	t-ratio	p-value
Intercept	483.06	10.81	44.70	<.001**
Level 1 Predictor:				
Home study support	4.25	1.15	3.69	<.001**
Level 2 Predictors:				
Model for school mean (π_0):				
The topic 'Numbers' taught to TIMSS classes	2.03	1.21	1.68	.102
The topic 'Algebra' taught to TIMSS classes	3.36	1.68	1.99	.054
Teachers' BS degree	1.41	.87	1.63	.113
Working load of teachers	-4.78	1.08	-4.44	<.001**
Number of students in the classes	.95	.28	3.35	.002**
Homework frequency	4.83	1.58	3.06	.004**
Teachers' preparedness to teach math,	1.16	1.19	.98	.336
Math homework usage index	1.37	.87	1.56	.127
Slope for home study support (π_1):				
Number of students in the classes	-.13	.052	-2.44	.020*
Teachers' BS degree	.87	.37	2.33	.026*

*Statistically significant at p<.05.
** Statistically significant at p <.01
Note: 46 iterations, 191 estimated parameters, deviance 2504048.67.

As explained at previous model (Model 2) home study support is a significant variable on mathematics achievement scores. As a contribution in model 3, school level effects are considered. As expected, results suggest that when working load of teachers in a school gets higher the mathematics achievement scores are getting lower (-4.78). Additionally, it seems that when mathematics homework frequency gets higher, the mathematics achievement scores become also higher (4.83). Contrary to what expected, results suggest that when the average number of students in classes gets higher, the mathematics achievement scores become higher (0.95) too. In other words, achievement levels get higher for schools which have crowded classes.

On the other hand, when intersection effects are considered, students who have more home study support and attend to schools which have less number of students in their classes get higher mathematics achievement scores (-0.13). In other

words, if the students who have higher home study support attend to the schools which have less crowded classes can have higher mathematics achievement scores.

Moreover, students who have more home study support and attend to schools which have teachers with high BS degrees which means have major both in mathematics and mathematics education or only mathematics education are able to get higher mathematics achievement scores (0.87). This is to say home study support and teachers who have major at education area ensure higher mathematics achievement.

The final estimation of random effects for model 3 which includes student level and school level predictor variables as conditional and country level predictor variables as unconditional are presented in Table 17, based on HLM3. According to the Table 17, the level-2 variance component is found to be significant, and all the level-3 variance components, except u_{06}, u_{14}, u_{15} and u_{16}, are found significant at the 0.05 level of significance level.

The explained variance at model 3 can be calculated as comparing the estimated variance at unconditional model and model 3 for level-1 and level-2.

$$\text{The explained variance at level-1}=\frac{\sigma^2(\text{Model 1})-\sigma^2(\text{Model3})}{\sigma^2(\text{Model 1})}$$

$$=\frac{4788.60-4770.34}{4788.60}=0.003$$

$$\text{The explained variance at level-2}=\frac{\tau_\pi(\text{Model 1})-\tau_\pi(\text{Model3})}{\tau_\pi(\text{Model 1})}$$

$$=\frac{2655.68-2302.89}{2655.68}=0.133$$

Table 17. HLM3 Analysis Results of Model 3-Random Effects (Level-1 and Level-2 Variables Included)

Random Effects	SD	Variance Component	d.f.	X^2	p-value
Intercept,r_0	47.98851	2302.89690	6105	101187.8729	<0.001**
level-1, e	69.06762	4770.33652			
INTRCPT1/INTRCPT2,u_{00}	64.72578	4189.42606	34	10351.79714	<0.001**
INTRCPT1/ HWFR,u_{01}	7.94974	63.19837	34	118.24423	<0.001**
INTRCPT1/ STUD,u_{02}	1.59873	2.55592	34	400.80696	<0.001**
INTRCPT1/ NUMB,u_{03}	5.31058	28.20229	34	85.49334	<0.001**
INTRCPT1/ ALGB,u_{04}	8.28784	68.68836	34	56.12848	0.010*
INTRCPT1/ PREP,u_{05}	5.26230	27.69185	34	73.76490	<0.001**
INTRCPT1/ HMWC,u_{06}	3.10713	9.65427	34	35.92694	0.378
INTRCPT1/ LOAD,u_{07}	4.78895	22.93403	34	83.62794	<0.001**
INTRCPT1/ DEGR,u_{08}	3.52132	12.39969	34	62.60409	0.002**
HSUP/INTRCPT2,u_{10}	6.60555	43.63324	34	487.05965	<0.001**
HSUP/ HWFR,u_{11}	1.54977	2.40178	34	51.66592	0.026*
HSUP/ STUD,u_{12}	0.20355	0.04143	34	50.78802	0.032*
HSUP/ NUMB,u_{13}	1.98487	3.93972	34	69.22169	<0.001**
HSUP/ ALGB,u_{14}	1.73358	3.00528	34	34.17946	0.459
HSUP/ PREP,u_{15}	1.29705	1.68233	34	33.69844	>.500
HSUP/ HMWC,u_{16}	0.96746	0.93597	34	27.84597	>.500
HSUP/ LOAD,u_{17}	1.77420	3.14779	34	62.44737	0.002**
HSUP/ DEGR,u_{18}	1.38010	1.90469	34	49.72606	0.040*

*Statistically significant at p<.05.
** Statistically significant at p <.01

According to the result of explained variance at level-1, approximately 0.3% of the variance of between countries mathematics achievement scores is explained by the predictor variable “home study support”. It is obtained at unconditional model that student level explains 41% of mathematics achievement scores, so it can be stated that approximately 0.1% (0.41*0.003=0.001) of the variance at mathematics achievement is sourced from the predictor variable “home study support”, when the mentioned school level predictors are controlled and country level predictors are not controlled. Furthermore, according to the result of explained variance at level-2, approximately 13.3% of the variance of between countries mathematics achievement scores is explained by the predictor variables homework frequency, number of students in the classes, the topics ‘Numbers’ taught to TIMSS classes, the topics

'Algebra' taught to TIMSS classes, preparation to teach mathematics, mathematics homework for the TIMSS class, working load of teachers and teachers' BS degree. It is obtained at unconditional model that school level explains 23% of mathematics achievement scores, so it can be stated that approximately 3% (0.23*0.133=0.031) of the variance at mathematics achievement is sourced from the mentioned predictor variables, when the student level predictor variable home study support is controlled and country level predictors are not controlled. The estimations of reliabilities for estimating student and school level model outcomes are given in Table 18.

Table 18. Final Estimation of Level-1 and Level-2 Reliabilities

Coefficient	Reliability estimate
Level 1 Intercept, π_0	0.928
Level 2 Intercept, β_{00}	0.996
Intercept1/ HWFR, β_{01}	0.677
Intercept1/ STUD, β_{02}	0.857
Intercept1/ NUMB, β_{03}	0.510
Intercept1/ ALGB, β_{04}	0.659
Intercept1/ PREP, β_{05}	0.516
Intercept1/ HMWC, β_{06}	0.333
Intercept1/ LOAD, β_{07}	0.536
Intercept1/ DEGR, β_{08}	0.444
HSUP/Intercept2, β_{10}	0.903
HSUP/ HWFR, β_{11}	0.306
HSUP/ STUD, β_{12}	0.352
HSUP/ NUMB, β_{13}	0.393
HSUP/ ALGB, β_{14}	0.293
HSUP/ PREP, β_{15}	0.248
HSUP/ HMWC, β_{16}	0.190
HSUP/ LOAD, β_{17}	0.400
HSUP/ DEGR, β_{18}	0.352

In sum, the findings of significant ($p<.01$) intercepts in school level suggest the development of the model as specifying country-level variables as extending the model 3 in order to formulate the final model. With the intention of adding level-3 predictor variables into the model, parallel to research question 2, the same predictor variables "components prescribed by mathematics curriculum" and "instructional time" are included to the final model. Therefore, for final model analysis, two

curriculum variables (COMP and INST), eighth school level predictor variables (DEGR, LOAD, STUD, NUMB, ALGB, PREP, HMWC and HWFR) and one student level predictor variable (HSUP) are tested.

4.7 Results for final model

In this final step of the research, student, school and country levels are all taken as conditional and formulated in order to examine the overall effect of the predictor variables on mathematics achievement, when student, school and country level predictor variables considered together. As mentioned at previous section, one student-level predictor variable (home study support), eighth school level predictor variables (homework frequency, number of students in the classes, the topics 'Numbers' taught to TIMSS classes, the topics 'Algebra' taught to TIMSS classes, preparation to teach mathematics, mathematics homework for the TIMSS class, working load of teachers and teachers' BS degree) and two level-3 predictor variables (components prescribed by mathematics curriculum and instructional time) are considered while performing analysis for final model.

The final estimation of fixed effects for final model is presented in Table 19, based on HLM3. As seen on Table 19, final model is more complex than model 3 which has only Level-1 and Level-2 predictor variables as controlled. Firstly, mathematics achievement scores are directly related to the student characteristic, home study support. Secondly, effects of this variable are moderated by school variable, teachers' BS degree. In addition, four out of eighth school characteristics are directly related to mathematics achievement scores similar with the previous model and three out of eighth school characteristics moderated by country level

predictor variables. Finally, two country characteristics are directly related to mathematics achievement scores.

Table 19. HLM3 Analysis Results of Model 4 (Level-1, Level-2 and Level-3 Variables Included)

Fixed Effect	Coeffi-cient	SE	T-ratio	p-value
Intercept	483.06	9.52	50.71	<.001**
Level 1 Predictor:				
Home study support	4.19	1.06	3.95	<.001**
Level 2 Predictors:				
Model for school mean (π_0):				
Teachers' BS degree	1.26	.80	1.57	.126
Topics 'Algebra' taught to TIMSS classes	3.48	1.63	2.13	.040*
Working load of teachers	-4.56	.94	-4.86	<.001**
Number of students in the class	.97	.26	3.78	<.001**
Homework frequency	5.70	1.50	3.79	<.001**
Slope for home study support (π_1)				
Teachers' BS degree	.77	.37	2.06	.047*
Level 3 Predictors:				
Model for country mean (β_{00}):				
Components prescribed by mathematics curriculum ***	-18.42	9.76	-1.89	.068 .038*
Instructional time	-1.05	.41	-2.56	.015*
Slope for homework frequency (β_{01}):				
Components prescribed by mathematics curriculum	-4.11	1.53	-2.68	.011*
Slope for number of student in classes (β_{02}):				
Components prescribed by mathematics curriculum	-.77	.27	-2.89	.007**
Slope for working load of teachers (β_{03}):				
Instructional time	-.10	.04	-2.69	.011*
Slope for teachers' BS degree (β_{04}):				
Instructional time	.08	.03	2.51	.017*
Slope for home study support (β_{10}):				
Components prescribed by mathematics curriculum	-2.49	1.10	-2.26	.031*

*Statistically significant at p<.05.
** Statistically significant at p <.01
***p-value with robust standard errors is p<.05.
Note: 51 iterations, 64 estimated parameters, deviance 2504103.30.

As explained at previous models "home study support" and "homework frequency", "number of students in the classes", "working load of teachers" and "the

topics ‘Algebra’ taught to TIMSS classes” are significant variables on mathematics achievement scores. Additionally, the variable “teachers’ BS degree” is also taken into the models in order to detect the interaction effects. As a contribution in final model, country level effects are considered. In this model, only the results which are different than previous models are presented.

As expected, the direct effects of the country level predictor variables are appeared very similar with the results of the model 1a (Research question 2). The coefficient of “components prescribed by mathematics curriculum” (γ_{002}) is estimated as -18.42 with standard error 9.76, which meant more centralized education systems cause lower mathematics achievement score average. The effect of countries’ “components prescribed by mathematics curriculum” index on the average mathematics achievement scores is statistically significant at $p<.05$ ($p=.038$) level with robust standard error. Furthermore, the coefficient of “instructional time” (γ_{001}) is estimated as -1.05 with standard error 0.41, which means spending more time on mathematics lessons is related with mathematics achievement score average negatively at a small amount. The effect of countries’ “instructional time” on the average mathematics achievement scores is statistically significant at $p<.05$ level.

When the effects of country level variables on school level slopes are examined, results about the curriculum level effects suggest that homework frequency and number of students in the classes are especially important related with the amount of components prescribed by mathematics curriculum. If the amount of components prescribed by curriculum is less, which means less centralized systems, in order to get higher mathematics achievement, homework should be assigned more often. According to this finding, it can be stated that because the system gives the

responsibility to more local administrations in less centralized systems, homework can be used as a check mechanism.

Furthermore, if the index of components prescribed by mathematics curriculum is higher, which means more centralized systems, in order to get higher mathematics achievement, number of students in the classes should be less. This finding, support less crowded classrooms in more centralized education systems.

In addition, for the effects of country level variables on school level slopes, results about the curriculum level effects suggest that teacher' BS degree and working load are especially important related with the instructional time specified in mathematics curriculum. For instance, if the instructional time gets high, which means busier schedule for the teacher, in order to get higher mathematics achievement for students, working load of teachers about other things than instruction should be less. On the other hand, if teachers' BS degree is high, than instructional time should be higher to get higher mathematics achievement.

Some country level characteristics are related with student level slopes. Countries which have less centralized educational system seem to be more successful in getting higher mathematics achievement scores of students' who have higher home study support. This shows that centralization of the system decreased the effect of social or economic differences. The final estimations of random effects for final model are presented in Table 20, based on HLM3.

According to the Table 20, the level-2 variance component is found to be significant at 0.001 level of significance, and all the level-3 variance components, except u_{11} and u_{12} are found to be significant at the 0.05 level of significance.

Table 20. HLM3 Analysis Results of Final Model-Random Effects (Level-1, Level-2 and Level-3 Variables Included)

Random Effects	SD	Variance Component	d.f.	X^2	p-value
INTRCPT1,r_0	48.80748	2382.16987	6059	104423.7146	<0.001**
HSUP slope,r_1	8.44253	71.27624	6129	6956.28311	<0.001**
level-1, e	68.93434	4751.94372			
INTRCPT1/INTRCPT2,u_{00}	57.00854	3249.97332	32	7155.82861	<0.001**
INTRCPT1/ HWFR,u_{01}	7.24620	52.50735	32	101.78340	<0.001**
INTRCPT1/ STUD,u_{02}	1.41271	1.99575	32	274.40737	<0.001**
INTRCPT1/ LOAD,u_{03}	3.47362	12.06603	32	58.14293	0.003**
INTRCPT1/ DEGR_R,u_{04}	2.78859	7.77621	32	48.46812	0.031*
HSUP/INTRCPT2,u_{10}	5.95330	35.44181	32	377.57374	<0.001**
HSUP/ STUD,u_{11}	0.14318	0.02050	32	40.49964	0.144
HSUP/ DEGR_R,u_{12}	0.83742	0.70127	32	40.85756	0.136

*Statistically significant at p<.05.
** Statistically significant at p <.01

For the final model explained variances at each level can be calculated as comparing with unconditional model.

$$\text{The explained variance at level-1}=\frac{\sigma^2(\text{Model 1})-\sigma^2(\text{Model4})}{\sigma^2(\text{Model 1})}$$

$$=\frac{4788.60-4751.94}{4788.60}=0.008$$

$$\text{The explained variance at level-2}=\frac{\tau_\pi(\text{Model 1})-\tau_\pi(\text{Model4})}{\tau_\pi(\text{Model 1})}$$

$$=\frac{2655.68-2382.17}{2655.68}=0.103$$

$$\text{The explained variance at level-3}=\frac{\tau_\beta(\text{Model 1})-\tau_\beta(\text{Model4})}{\tau_\beta(\text{Model 1})}$$

$$=\frac{4195.69-3249.97}{4195.69}=0.225$$

All the variables from three levels were included into the final model at the same time. The results displayed that the student level variable, home study support, explained only 0.8% of the total student level variance when the school and country level variables are taken into analysis as conditional. The school level variables

explained 10.3% of the school variance when both student and country level variables are taken into analysis as conditional. The country level variables explained 22.5% of the country level variance, when both student and school level variables are taken into analysis as conditional. The explained variance for country level factors (22.5%) gives almost same variance as the Model 1a which did not control student and school level factors. In other words, controlling the student and school level variables did not make a difference for the amount of country level explained variance.

The estimations of reliabilities for estimating student and school level model outcomes are given in Table 21. According to the Table 21, the reliability estimates of level-1 and level-2 intercepts are given reasonably high.

Table 21. Final Estimation of Level-1 and Level-2 Reliabilities

Coefficient	Reliability estimate
Level 1 Intercept, π_0	0.932
HSUP,π_1	0.125
Level 2 Intercept, β_{00}	0.995
Intercept1/ HWFR, β_{01}	0.645
Intercept1/ STUD, β_{02}	0.831
Intercept1/ LOAD, β_{03}	0.389
Intercept1/ DEGR, β_{04}	0.340
HSUP/Intercept2, β_{10}	0.874
HSUP/ STUD, β_{11}	0.231
HSUP/ DEGR, β_{12}	0.167

CHAPTER 5

DISCUSSION

In this chapter, findings of the current study are discussed as considering the related literature for extensive understanding of the research and its findings. This chapter also includes limitations of the study and recommendations for further research.

5.1 The explained variance in different levels of hierarchical model

The purpose of this research is to determine whether Trends in International Mathematics and Science Study (TIMSS) 2015 eighth grade mathematics achievement scores vary between countries. If there is a variation, it is aimed to figure out the source of this variance at curriculum, school and student levels and in this way figure out the relation between intended, implemented and attained curricula. The findings showed that mathematics achievement scores varied significantly at all three levels. When unconditional model which does not control any level predictor variable was considered, 41.1% of total variance was derived from student level, 22.8% was derived from school level and 36.0% was derived from country level differences. It is presented that the amount of explained variance is highest at student level, which is followed by country level and lowest at school level. These findings pointed out the fact that country level factors have greater effects on students' mathematics achievement scores than school practices.

This sequence of variance is in line with some findings from literature. For instance, Çelik (2016) stated that total variance was shared out at student, school and country levels as 42.8%, 24.9% and 32.3% respectively. Moreover, Mohammadpour and Mohamed Najib (2014) and Ryoo (2001; as cited in Çelik 2016) reported similar

variance proportions, which demonstrate that the highest variances were sourced by within-school differences and lowest variances were sourced by between-school variances.

Furthermore, explained variances for each model were calculated. When only level-1 predictor variable is considered, it is found that 'home study support' explained only 0.3% of total variance in mathematics achievement scores of countries, whereas 3% of total variance in mathematics achievement scores of countries is explained by the considered eighth predictor variables at school level and 8% of total variance in mathematics achievement scores of countries is explained by the curriculum variables of "components prescribed by mathematics curriculum" and "instructional time".

On the other hand, when different levels of variables are controlled at the same time, it is found that the highest proportion of variance belong to country level with the curriculum variables "components prescribed by mathematics curriculum" and "instructional time", which means that country level policy decisions have bigger effect than lower level effects.

5.2 The influence of curriculum variables on mathematics achievement

The focus of the current study was the influence of curriculum variables on mathematics achievement. Contrary to the literature, coverage of mathematics topics (Mitchell, 1999; Squires, 2012; Boscardin et al., 2005; Porter, 2002), evaluation ways of curriculum (Bishop & Woesmann, 2004; Bol et al., 2013), mathematical and numeracy skills at pre-primary education (Berlinski, Galiani & Gertler, 2009; Tucker-Drop, 2012) and classroom tracking by achievement (Lamb & Fullarton, 2002; Van de Werfhorst & Mijs, 2010) had non-significant effect according to the

exploratory analyses. On the other hand, insignificant result about evaluation ways is consistent with findings of Van de Werfhorst & Mijs (2010) or Ruiz-Primo et al. (2002), mathematical and numeracy skills at pre-primary education is consistent with findings of Van de Rijt et al. (2003) and classroom tracking by achievement is consistent with Hanushek & Wössmann (2006). Hence, only two curriculum variables could be included to the further analyses. These variables were "components prescribed by mathematics curriculum" index and "instructional time" variable.

"Components prescribed by mathematics curriculum" variable meant how much detailed the curriculum stated by a central authority. Higher scores from this variable indicate more centralized systems. According to the findings, when a country's "components prescribed by mathematics curriculum" index increased one standard point, which indicated a more centralized system, the average mathematics achievement point of that country decreased 18 points, which is a very important difference. In other words, centralization of an education system is negatively related with mathematics achievement scores critically.

As Rado (2010) mentioned off course "high-quality policymaking and implementation is not a simple question of the capacities of the participating actors" (p.318), but according to the findings of this current study, decentralization is supported for getting higher student achievement levels which refers educational effectiveness, parallel to Rado (2010)'s arguments. Rado (2010) suggests transferring the power from the center to lower levels of administration in order to make the educational quality higher. Because centralized systems have hardness to adjust their curricula according to the needed adjustments on a small scale, they might have implementation problems. Bishop and Woesmann (2004) investigated the

administrative features of educational systems and school systems from various perspectives. They stated that in educational processes, students and teachers made decisions in order to maximize their own profit which can be higher student outcomes, whereas governments or central administrations made decisions in order to minimize the cost. Hence, this might give a view of how the perspectives of decisions can change closer and further of central administrations. Whereas the decision of what is more important for a country belongs to the government of the countries, Bishop and Woesmann (2004) stated as "institutional policies may be much more promising to increase the quality of schooling than resources policies" (p.26), which supports the view of more decentralized curricula in order to get higher quality.

According to the findings sparing more time to mathematical lessons in curriculum might have a negative effect on mathematics achievements scores. This finding was emerged differently from expectation, where it is suggested to get higher achievement when students have more time on a particular topic (Berlinger, 1990). Thus, this finding might support diversified findings about instructional time in literature (Heyneman & Loxley, 1983; Gettinger, 1984; Berliner, 1990). The reason explained for the inexplicit results might be stated with the difference between the allocated time which is the intended time (mentioned time with this variable) and active learning time of students (time-on-task) (Berliner, 1990; Cattaneo, Oggenfuss & Wolter, 2016). Carroll (1989) mentioned that allocated time defines the given opportunity to learn for students, but allocated time might not sufficiently represent the amount of classroom applications. Therefore, the unexpected finding of the current study and the differential findings in literature might be sourced from the

inconsistency of the planned and applied timing. On the other hand, the interaction effects of instructional time and other level variables show more rational results.

5.3 The influence of interaction effects on mathematics achievement

The other aim of this study is to explain the links between students' mathematics achievements and possible predictor variables of achievement at student, school and country levels as considering the interaction effects. In order to make improvements after policy changes and national or regional curriculum changes, the applications in classroom and school environments have to be improved. Teachers are the implementers of curriculum on a specific content area in the classroom environments. Hence, teachers characteristics, such as their work load, preparedness, degree should also been considered together with students' economic status when curriculum implications in classroom environments desired to be considered. With that mentioned aim, the influence of the interaction of variables at student, school and country levels on mathematics achievement is considered as crucial as part of this current study. This part is discussed under three sub-titles as direct effects of each level predictor, interactions between two levels and interactions between three levels together.

5.3.1 The direct interaction effects

Firstly, more home study support on student level had a positive relation with students' mathematics achievement scores. Because home study support was included into the study as an indicator of students' socio-economic status, this finding can be considered as parallel with the literature which supports the positive

relation between mathematics achievement and socio-economic status (e.g. Yang, 2003; O'Dwyer, 2005; Chiu, 2010).

Secondly, when school level classroom context variables are considered, results displayed that mathematic homework frequency and number of students in the classes had positive and working load of teachers had negative relations with mathematics achievement. More explicitly, parallel with the literature when teachers assigned homework more frequently mathematics achievement level became higher (Lamb & Fullarton, 2002; USDE, 1986). On the other hand, it is found that when class size became bigger, mathematics achievement increased, whereas many researches in literature suggested a negative relation (Summers & Wolfe, 1975; Eberts, Kehoe & Stone, 1984) or insignificant relation between class size and achievement (Darling-Hammond, 2000; Atar, 2014). The positive relation between student number and mathematics achievement might be based on the relation between poverty and class size. Additionally, parallel to the literature when working load of teachers becomes decreased, students' achievement levels become higher (Firestone & Pennel, 1993; Goddard, Hoy & Hoy, 2000; Johnson, Kraft & Papay, 2011; Bascia & Rottmann, 2011). In other words, when working conditions of teachers become better, such as less teaching hours, enough time for preparation, less administrative tasks, they can be more effective in the classroom, hence students can achieve higher.

5.3.2 The interaction effects between two levels

Thirdly, when the interactions between student and school level variables are considered, the results suggested that when the number of students in classes became less and students' home study support became high than mathematics achievement

level turned into high too. This finding might be interpreted as when students from higher socio-economic families attend to less crowded classes can get more attention on learning environment and this can be concluded as higher achievement. The findings of Eberts, Kehoe and Stone (1984)'s study to investigate the effect of school size together with other school factors are parallel with this finding of the current study. The study showed that differences in resources between students had a large effect on achievement. Socio-economic status or teachers' attention per student related with class size might be given as examples for this finding.

Another finding about this interaction showed that when both students' home study support and teachers' BS degrees increased than students' mathematics achievement also evolved into higher. Teachers' BS degree did not show significant relation solo with mathematics achievement, whereas the effect gets significant when considered with home study support, which shows that teachers' BS degree becomes particularly important for students' with more home study support.

5.3.3 The interaction effects between three levels

Last but not least, the interactions of variables in student, school and country levels were considered all together. According to the findings, homework frequency and number of students in classrooms became more important according to whether a country's curriculum elements were defined more centralized or decentralized. In less centralized systems, which curriculum elements are not decided strictly (e.g. for materials or evaluation methods) and the responsibility is assigned on schools more might get higher mathematics score as signing homework more frequently or as forming less crowded classes.

Additionally, working load of teachers and teachers' BS degree became more critical according to instructional time defined in each country's curriculum. When instructional time gets higher, in order to get higher mathematics achievement, it is needed to decrease teachers' working load. On the other hand, if teacher's BS degree is higher, which means if teacher has not only mathematics but also education background, than teacher can handle with higher instructional time and even make mathematics achievement scores higher as dealing with higher amount of assigned hours.

5.4 Limitations of the study

This research is limited with;

1. The determined questions of TIMSS 2015 student, teacher and curriculum questionnaires
2. TIMSS 2015 attendant countries' 8^{th} grade students and schools, except benchmarking countries
3. The aggregation of each variable at classroom and teacher level into school level which causes the ignorance of variation in between classroom in the same school
4. Lack of other excluded country level characteristics
5. The assumptions based on HLM approach, such as;

   5a. Errors have a normal distribution with "0" mean and errors are independent from each other with equal variance

   5b. Each independent variable is unrelated with the errors of its own level and the errors of other levels.

5.5 Recommendations and suggestions for further research

This study is one of the rare studies in the area which investigate the student, school and country level predictor variables simultaneously in the context of student, classroom and curriculum effects. Especially the revealed significant interaction effects of variables which do not have significant effect alone show the necessity of investigation of the interaction effects of different factors in education area.

The focus of this study as a contribution to the literature was curriculum factors, but there had been several limitations resulted from using an already existing data set. In order to get deeper perspective of curriculum affect, in larger quantities of data are needed. Hence, it might be useful to gather curriculum data from smaller units in countries or regions and investigate the curriculum affect better.

At last but not least, there were some unexpected findings revealed out of analyses. For instance, it was found that class size and mathematics achievement had positive relation and assigned instructional time on curriculum and mathematics achievement had negative relation. These variables needed to be examined with further researches. In addition, further researches may inspect the relation of curriculum factors and different disciplines such as science performance which related to mathematics and literature performance which differs with mathematics.

APPENDIX A

CORRELATION ANALYSES

Table A1. Correlations Between Level-1 Predictor and Dependent Variables

	Number of Home Study Support	1st PLAUSIBLE VALUE MATHEMATICS
Number of Home Study Support	1	.257**
1st PLAUSIBLE VALUE MATHEMATICS	.257**	1

** Correlation is significant at the 0.01 level (2-tailed).

Table A2. Correlations Between Level-2 Predictor Variables

	HWFR	STUD	TIME	NUMB	ALGB	GEOM	DATA	PREP	HMWC	LOAD	DEGR
HWFR	1	-.009	.140**	.168**	.114**	-.025*	-.040**	.068**	.208**	-.080**	-.006
STUD	-.009	1	.074**	.026*	.054**	.044**	-.056**	-.091**	.016	.192**	-.061**
TIME	.140**	.074**	1	.061**	.066**	.087**	.061**	.097**	.115**	-.018	-.044**
NUMB	.168**	.026*	.061**	1	.192**	.128**	.060**	.063**	.109**	-.053**	-.013
ALGB	.114**	.054**	.066**	.192**	1	.330**	.261**	.154**	.165**	.038**	.047**
GEOM	-.025*	.044**	.087**	.128**	.330**	1	.333**	.164**	.106**	.046**	.063**
DATA	-.040**	-.056**	.061**	.060**	.261**	.333**	1	.307**	.049**	.015	.062**
PREP	.068**	-.091**	.097**	.063**	.154**	.164**	.307**	1	.033**	.028*	.033**
HMWC	.208**	.016	.115**	.109**	.165**	.106**	.049**	.033**	1	-.045**	.034**
LOAD	-.080**	.192**	-.018	-.053**	.038**	.046**	.015	.028*	-.045**	1	-.015
DEGR	-.006	-.061**	-.044**	-.013	.047**	.063**	.062**	.033**	.034**	-.015	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table A3. Correlations Between Level-3 Predictor Variables

	COMP	INST	NTOP	EVAL	MPPE	CGRP
COMP	1	-.002	-.216	.247	-.041	.411*
INST	-.002	1	-.106	-.128	-.032	.185
NTOP	-.216	-.106	1	-.017	.107	-.192
EVAL	.247	-.128	-.017	1	.000	.103
MPPE	-.041	-.032	.107	.000	1	-.058
CGRP	.411*	.185	-.192	.103	-.058	1

* Correlation is significant at the 0.05 level (2-tailed).

APPENDIX B

RESIDUAL FILE ANALYSES

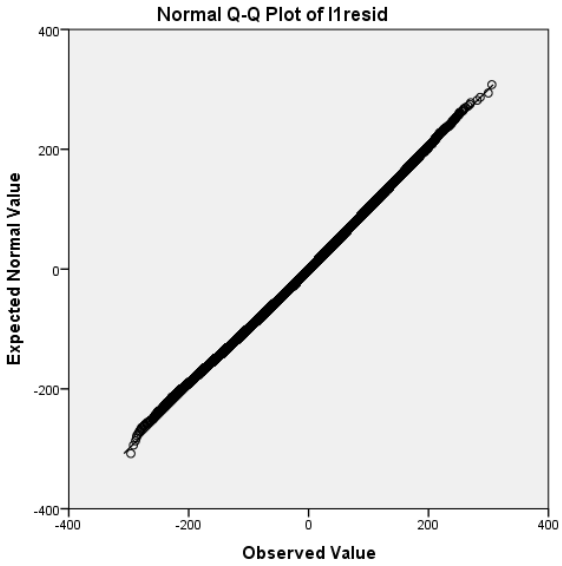


Figure B1. Normal Q-Q curve of Level-1 residuals for 219,788 students based on the final fitted model

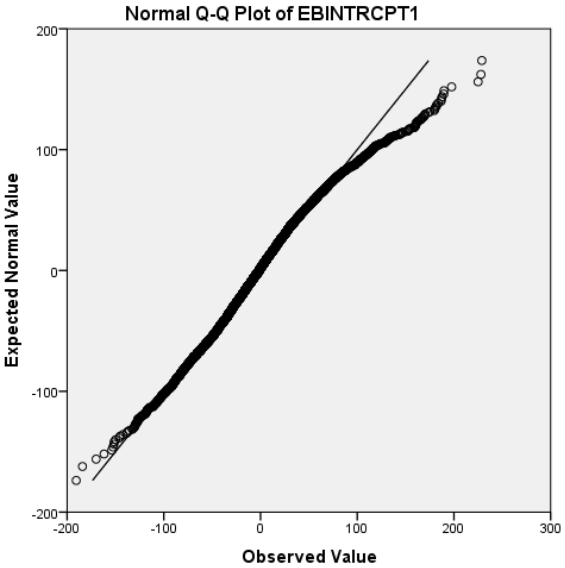


Figure B2. Normal Q-Q curve of Level-2 residuals for 6,586 schools based on the interception model

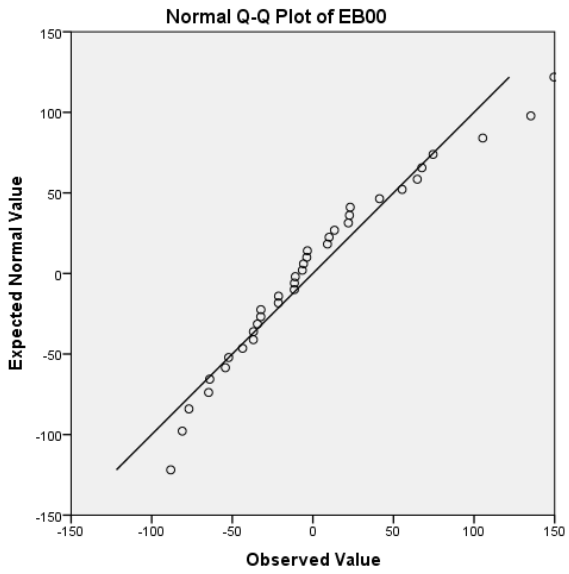


Figure B3. Normal Q-Q curve of Level-3 residuals for 36 countries based on the final fitted model

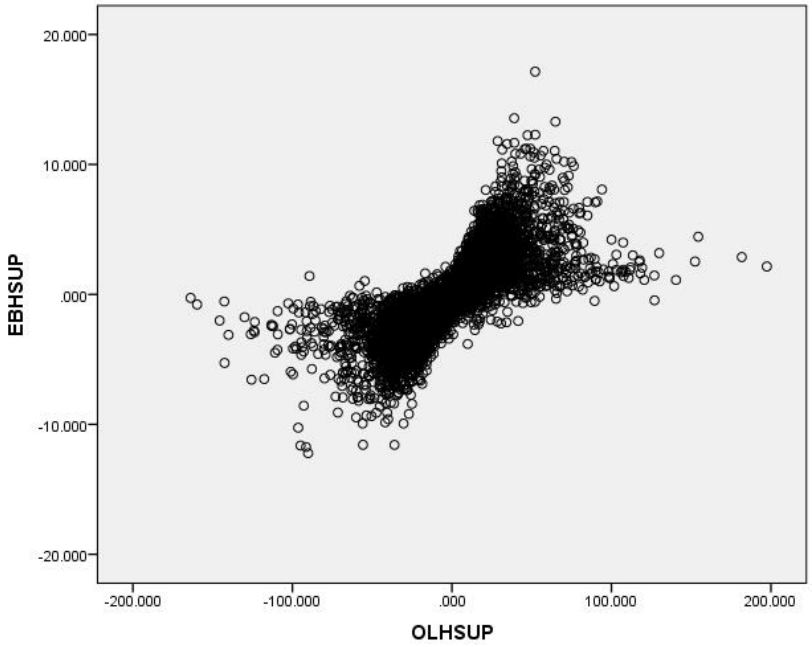


Figure B4. EB versus OL residuals for the home study support (HSUP) slopes in Level-2

APPENDIX C

DESCRIPTIVE STATISTICS FOR RESIDUAL FILES

Table C1. Descriptive Statistics of Level-2 Residual File

	N	Minimum	Maximum	Mean	Std. Deviation
NJK	6586	1	258	33.37	15.638
EBINTRCPT1	6586	-190.550	228.794	-.00002	46.552833
EBHSUP	6586	-12.219	17.142	-.00002	2.915620
OLINTRCPT1	6400	-225.115	249.505	-.51635	50.417915
OLHSUP	6400	-163.835	197.422	-.48369	29.841434
FVINTRCPT1	6586	289.445	696.626	475.89234	67.065606
FVHSUP	6586	-6.438	18.685	3.86873	6.373090
ECINTRCPT1	6586	258.341	734.638	475.89232	82.155945
ECHSUP	6586	-18.003	29.271	3.86871	7.054630
PV2_0_0	6586	50.902	1618.851	204.73302	105.815089
PV2_1_0	6586	-29.517	-.726	-3.38988	1.947298
PV2_1_1	6586	39.867	68.009	59.75922	4.860579
HWFR	6586	-3.387	2.472	.00000	.808181
STUD	6586	-36.612	53.603	.00000	8.128416
ALGB	6586	-3.281	2.290	.00000	.764712
LOAD	6586	-3.417	2.913	.00000	.931568
INST	6586	-38.444	56.556	3.10202	23.534341
Valid N (listwise)	6400				

Table C2. Descriptive Statistics of Level-3 Residual File

	N	Minimum	Maximum	Mean	Std. Deviation
NK	36	48	437	182.94	72.647
EB00	36	-100.092	125.126	.00005	61.445788
EB01	36	-14.029	19.711	.00071	7.186353
EB02	36	-2.775	4.196	-.00023	1.544612
EB03	36	-11.145	20.509	.00608	5.781706
EB04	36	-9.229	7.505	.00011	4.042780
EB10	36	-10.349	14.773	.00056	6.375256
OL00	35	-100.480	125.594	-3.58374	58.659205
OL01	35	-21.216	37.875	.85644	11.771831
OL02	35	-3.906	5.075	.02518	1.820607
OL03	35	-14.964	13.491	-1.38103	6.986101
OL04	35	-14.556	9.653	.18201	6.101867
OL10	35	-10.947	25.133	.24262	7.832903
FV00	36	448.630	506.482	483.07082	14.300348
FV01	36	5.398	5.398	5.39790	.000000
FV02	36	.932	.932	.93187	.000000

	N	Minimum	Maximum	Mean	Std. Deviation
FV03	36	3.734	3.734	3.73378	.000000
FV04	36	-4.776	-4.776	-4.77584	.000000
FV10	36	3.911	3.911	3.91131	.000000
EC00	36	373.193	613.071	483.07087	65.497004
EC01	36	-8.631	25.109	5.39861	7.186353
EC02	36	-1.844	5.128	.93165	1.544612
EC03	36	-7.411	24.242	3.73986	5.781706
EC04	36	-14.005	2.729	-4.77573	4.042780
EC10	36	-6.438	18.685	3.91187	6.375256
PV3_0_0_0_0	36	5.757	50.124	16.01944	7.195012
PV3_0_1_0_0	36	-.994	-.042	-.28944	.189717
PV3_0_1_0_1	36	6.280	31.253	15.53717	6.582481
PV3_0_2_0_0	36	-.117	.234	-.00147	.052047
PV3_0_2_0_1	36	.040	2.257	.58651	.517063
PV3_0_2_0_2	36	.039	.890	.28706	.186454
PV3_0_3_0_0	36	.029	.965	.24950	.181270
PV3_0_3_0_1	36	-5.773	.716	-.69235	1.264444
PV3_0_3_0_2	36	-1.470	.196	-.13472	.296841
PV3_0_3_0_3	36	7.129	35.817	18.54687	7.546570
PV3_0_4_0_0	36	-.339	.129	-.11255	.074348
PV3_0_4_0_1	36	-2.860	.085	-1.08045	.613377
PV3_0_4_0_2	36	-.016	1.334	.30316	.304580
PV3_0_4_0_3	36	-.252	2.015	.65456	.569926
PV3_0_4_0_4	36	3.032	10.858	6.28082	1.699152
PV3_1_0_0_0	36	-.614	-.049	-.20279	.090232
PV3_1_0_0_1	36	.121	3.115	.88790	.721929
PV3_1_0_0_2	36	-.388	-.009	-.09238	.079185
PV3_1_0_0_3	36	-.027	1.016	.29859	.267822
PV3_1_0_0_4	36	.176	3.578	1.10875	.747336
PV3_1_0_1_0	36	1.061	14.039	4.12963	2.728635
INST	36	-38.444	56.556	.00000	23.483057
Valid N (listwise)	35				

Table C3. Descriptive Statistics of Level-3 Residual File (Final Model)

	N	Minimum	Maximum	Mean	Std. Deviation
NK	36	48	437	182.94	72.647
EB00	36	-88.202	149.621	.00011	57.643797
OL00	36	-88.505	150.287	.01250	57.929836
FV00	36	423.000	551.711	483.19541	31.219044
EC00	36	373.190	612.820	483.19553	65.554913
PV3_0_0_0_0	36	6.411	55.931	17.82238	8.028371
INST	36	-38.444	56.556	.00000	23.483057
COMP	36	-1.505	.784	.00000	.990225
Valid N (listwise)	36				

APPENDIX D

EXPLORATORY ANAYSES

Table D1. Exploratory Analysis for Estimated Level-2 Coefficients and Standard Errors

Level-1 Coefficient	Potential Level-2 Predictors					
INTRCPT1,π_0						
	HWFR	STUD	TIME	NUMB	ALGB	GEOM
Coefficient	4.922	0.145	-0.005	2.649	2.643	0.492
Standard Error	0.756	0.075	0.011	0.701	0.801	0.718
t-value	6.511	1.918	-0.470	3.780	3.301	0.685
INTRCPT1,π_0						
	DATA	PREP	HMWC	LOAD	DEGR_R	
Coefficient	0.971	1.874	1.390	-4.678	2.422	
Standard Error	0.739	0.729	0.674	0.655	0.603	
t-value	1.313	2.569	2.063	-7.137	4.017	

Table D2. Exploratory Analysis for Estimated Level-3 Coefficients and Standard Errors

Level-1 Coefficient	Potential Level-3 Predictors					
INTRCPT1/INTRCPT2,β_{00}						
	INST	NTOP	MPPE	CGRP	COMP	EVAL
Coefficient	-1.072	0.913	15.938	-22.437	-19.030	-3.565
Standard Error	0.479	3.350	22.204	24.749	11.353	11.107
t-value	-2.239	0.272	0.718	-0.907	-1.676	-0.321

REFERENCES

Akyüz, G. (2006). Investigation of the effect of teacher and class characteristics on mathematics achievement in Turkey and European Union Countries. *Elementary Education Online,* 5(2), 75-86.

Anastasi, A. (1982). *Psychological testing* (5th edition). New York: Macmillan.

Arora, A. (2010). Relationship among curriculum coverage, teacher preparedness, and student achievement in TIMSS Advanced 2008. *TIMSS&PIRLS International Study Center*, Boston College.

Atar, H. Y. (2014). Öğretmen niteliklerinin TIMSS 2011 fen başarısına çok düzeyli etkileri. *Eğitim ve Bilim*, 39 (172), 121-137.

Bascia, N. & Rottmann, C. (2011). What's so important about teachers' working conditions? The fatal flaw in North American educational reform. *Journal of Education Policy,* 26(6), 787-802.

Berliner, D. C. (1990). What's all the fuss about instructional time? *The Nature of Time in Schools Theoretical Concepts, Practitioner Perceptions*. New York: Teachers College Press, Columbia University.

Berlinski, S., Galiani, S. & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93 (1-2), 219-234.

Bishop, J. & Woesmann, L. (2004). Institutional effects in a simple model of educational production. *Education Economics*, 12, 17-38. http://digitalcommons.ilr.cornell.edu/hrpubs/20/

Bol, T., Witschge, J., Van de Werfhorst, H. G. & Dronkers, J. (2013). Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Amsterdam Centre of Inequality Studies (AMCIS) Working Paper Series.*

Boscardin, C. K., Aguirre-Munoz, Z., Stoker, G., Kim, J., Kim, M. & Lee, J. (2005). Relationship between opportunity to learn and student performance on English and algebra assessments. *Educational Assessment,* 10(4).

Bray, M. (2005). Methodology and focus in comparative education. In M. Bray and R. Koo (Eds.), *Education and society in Hong Kong and Macao* (2nd ed.) (pp. 239-252). Netherlands: Springer.

Cal, G. (2011). *Opportunity to learn (OTL) and the alignment of upper division mathematics learning outcomes, textbooks, and the national assessment in Belize* (Unpublished doctoral dissertation). University of South Florida: Florida. UMI Number: 34481563.

Carroll, J.B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26-31.

Cattaneo, M.A., Oggenfuss, C. & Wolter, S.C. (2016). The more, the better? The impact of instructional time on student performance. *Discussion Paper Series*, No.9797.

Chepete, P.L. (2008). *Modeling of the factors affecting mathematical achievement of form 1 students in Botswana based on the 2003 trends in international mathematics and science study* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.

Chiu, M.M. (2010). Effects of inequality, family and school on mathematics achievement: Country and student differences. *Social Forces*, 88(4), 1645-1676.

Cogan, L.S. & Schmidt, W.H. (1999). An examination of instructional practices in six countries. In G. Kaiser, E. Luna & I. Huntley (Eds.) *International comparisons in mathematics education* (pp.68-85). London: Falmer Press.

Crocker, L. & Algina, J. (2008). *Introduction to classical & modern test theory*. Ohio: Cengage Learning.

Çelik, İ. (2016). *Ülke özelliklerinin TIMSS 2011 sekizinci sınıf matematik başarısına çok düzeyli etkileri* (Unpublished master's thesis). Gazi University, Ankara, Turkey.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1-44.

Eberts, R. W., Kehoe, E. & Stone, J. A. (1984). The effect of school size on student outcomes. *University of Oregon, Eugene. Center of Educational Policy and Management*. (ERIC Document Reproduction Service No. ED 245382).

Firestone, W. A. & Pennell, J. R. (1993). Teacher commitment, working conditions, and differential incentive policies. *Review of Educational Research*, 63 (4), 489-525.

Foy, P. (2017). *TIMSS 2015 User Guide for the International Database*. Boston: TIMSS & PIRLS International Study Center.

Foy, P., Martin, M. O., Mullis, I. V. S., Yin, L., Centurino, V. A. S., & Reynolds, K. A. (2016). Reviewing the TIMSS 2015 Achievement Item Statistics. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 11.1-11.43). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-11.html

Frankel, J. R., Wallen, N. E. & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Gay, L.R., Mills, G. E. & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Boston: Pearson Education, Inc.

Gettinger, M. (1984). Individual differences in time needed for learning: A review of literature. *Educational Psychologist*, 19(1), 15-29. DOI: 10.1080/00461528409529278

Goddard, R., Hoy, W.K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37(2), 479-507.

Grønmo, L. S., Lindquist, M, Arora, A & Mullis, I.V.S. (2013). In I.V.S. Mullis. & M. O. Martin (Eds.) (2013). *TIMSS 2015 Assessment Frameworks* (p.11-27). Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

Hanushek, E. A. & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in differences evidence across countries. *The Economic Journal*, 116, C63-C76.

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analysis relating to achievement*. London: Routledge.

Heyneman, S.P. & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology*, 88(6), 1162-1194.

Hill, C. H., Sleep, L., Lewis, M. J., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111-155). Charlotte, NC: Information Age Publishers.

Ho, E. S. (2009). Characteristics of East Asian learners: What we learned from PISA. *Educational Research Journal*, 24(2), 327-348.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23(6), 723-744.

Hofmann, D. A., Griffin, M. A. & Gavin, M. B. (2000). The application of hierarchical modeling to organizational research. In K.J. Klein & S.W.J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp.467-511). San Francisco, CA: Jossey-Bass.

Hooper, M. (2016). Developing the TIMSS 2015 context questionnaires. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015 (*pp. 2.1-2.8). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-2.html

Hooper, M., Mullis, I.V.S. & Martin, M.O. (2013). TIMSS 2015 context questionnaire framework. In I.V.S. Mullis, & M.O. Martin (Eds.), *TIMSS 2015 assessment frameworks.* Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation* (NBER Working Paper No. 7867). Cambridge, MA: National Bureau of Economic Research.

IEA: International association for the evaluation of educational achievement (2011). Retrieved 29 January, 2016, from http://www.iea.nl/?id=72

IEA: International association for the evaluation of educational achievement (2014). Teacher Questionnaire Mathematics-TIMSS 2015. TIMSS&PIRLS International Study Center, Lynch School of Eduaction, Boston College

ISCED: International standards classification of education ISCED 2011 (2012). Montreal: UNESCO Institute for Statistics. Retrieved 17 March, 2017, from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

Jacobs, H.H. (2010). A new essential curriculum for a new time. In H.H. Jacobs (Eds.), *Curriculum 21: essential education for a changing world* (pp.7-17). Alexandria, VA: ASCD.

Johansone, I. (2016). Survey Operations Procedures in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 6.1-6.22). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-6.html

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2011, June 29). How context matters in high-need schools: The effect of teachers' working conditions on their professional satisfaction and their students' achievement. *Project on the Next Generation of Teachers Harvard Graduate School of Education*. Retrieved from https://scholar.harvard.edu/files/mkraft/files/johnson_kraft_papay_teacher_working_conditions_final.pdf

Joncas, M. & Foy, P. (n.d). Sample design in TIMSS and PIRLS. In M. O. Martin & I. V. S. Mullis (eds.) *Methods and procedures in TIMSS & PIRLS 2011*. Boston College, Lynch School of Education. Retrieved from http://timssandpirls.bc.edu/methods/pdf/TP_Sampling_Design.pdf

Kreft, I.G.G., de Leeuw, J. & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30(1), 1-21.

Lamb, S. & Fullarton, S. (2002). Classroom and school factors affecting mathematics achievement: A comparative study of Australia and the United States using TIMSS. *Australian Journal of Education*, 46(2), 154-171.

LaRoche, S., Joncas, M., & Foy, P. (2016). Sample Design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 3.1-3.37). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-3.html

Maas, C. J. & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92.

Mandeville, G. K. & Liu, Q. (1997). The effect of teacher certification and task level on mathematics achievement. *Teaching and Teacher Education*, 13 (4), 397-407.

Martin, M.O., Mullis, I.V.S. & Foy, P. (2013). TIMSS 2015 assessment design. In I.V.S. Mullis & M.O. Martin (Eds.) (2013), *TIMSS 2015 assessment frameworks*. Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., & Palazzo, L. (2016). Creating and interpreting the TIMSS 2015 context questionnaire scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 15.1-15.312). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-15.html

McNamara, D. R. (1981). Attention, time-on-task and children's learning: research or ideology? *Journal of Education for Teaching: International Research and Pedagogy*, 7(3), 284-297, DOI: 10.1080/0260747810070308.

Mitchell, F.M. (1999). All students can learn: Effects of curriculum alignment on the mathematics achievement on third-grade students. Paper presented at: *American Educational Research Association Annual Meeting*. Montreal, Canada.

Mohammadpour, E. & Mohamed Najib, A.G. (2014). Mathematics achievement as a function of within- and between-school differences. *Scandinavian Journal of Educational Research*, 58(2), 189-221. Retrieved from http://dx.doi.org/10.1080/00313831.2012.725097

Mugo, P. & Wolhuter, C.C. (2013). Definition and purpose of comparative education. In C.C. Wolhuter, I. Kamere & K. Biraimah (Eds.), *A student's textbook in comparative education* (pp.1-21). Potchefstroom: Platinum Press.

Mullis, I.V.S. (2013). Introduction. In I.V.S. Mullis & M.O. Martin (eds.*) TIMSS 2015 Assessment Framework.* . TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Mullis, I. V. S., Cotter, K. E., Fishbein, B. G., & Centurino, V. A. S. (2016). Developing the TIMSS 2015 achievement items. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp.1.1-1.22). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-1.html

Mullis, I.V.S., Drucker, K.T., Preuschoff, C., Arora, A. & Stanco, G.M (2012). Assessment framework and instrument development. In M. O. Martin & I. V. S. Mullis (eds.) *Methods and procedures in TIMSS & PIRLS 2011.* Retrieved from http://sling11.bc.edu/methods/pdf/TP_Instrument_Devel.pdf

Mullis, I. V. S., Martin, M. O., Foy & Arora, A. (2012). *TIMSS international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Goh, S. & Cotter, K. (Eds.) (2016). *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.ed/timss2015/encyclopedia/

Mullis, I. V. S., Martin, M. O., Minnich, C. A., Stanco, G. M., Arora, A., Centurino, V. A. S. & Castle, C. E. (2012). *TIMSS 2011 encyclopedia education policy and curriculum in mathematics and science volume 1: A-K*. Chestnut Hill, MA: TIMSS&PIRLS International Study Center, Boston College.

Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. International Association for the Evaluation of Educational Achievement.

National Research Council (1996). *Mathematics and science education around the world: What can we learn?* National Academy of Sciences. Retrieved 03 February, 2017 from: http://www.nap.edu/catalog/9146.html

Newman, D., Newman, I. & Salzman, J. (2010). Comparing OLS and HLM models and the questions they answer: Potential concerns for type VI errors. *Multiple Linear Regression Viewpoints*, 36 (1), 1-8.

Null, W. (2011). *Curriculum: From theory to practice*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.

Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: Cambridge University Press.

O'Dwyer, L. M. (2005). Examining the variability of mathematics performance and its correlates using data from TIMSS '95 and TIMSS '99. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 11(2), 155-177. DOI: 10.1080/13803610500110802

OECD (2004). What makes school systems perform? Seeing school systems through the prism of PISA. Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33858946.pdf

Osborn, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation*, 7(1), 1-4.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, 30(1), 66-85.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.

Rado, P. (2010). *Governing decentralized education systems: Systemic change in South Eastern Europe*. Budapest, Hungary: Open Society Foundations.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2011). HLM 7.00 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S., Bryk, A., Cheong, Y.F., Congdon, R. & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S., Bryk, A., Cheong, Y.F., Congdon, R. & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Roseth, C. J., Johnson, D. W. & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin*, 134 (2), 223-246.

Ruiz-Primo, M.A. Shavelson, R.J., Hamilton, L. & Klein, S. (2002). On the evaluation of systematic science education reform: searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39(5), 369-393.

Schleicher, A. (2007). How the world's best-performing school systems come out on top. Retrieved from http://www.smhc-cpre.org/wp-content/uploads/2008/07/how-the-worlds-best-performing-school-systems-come-out-on-top-sept-072.pdf

Schütz, G., Ursprung, H. W. & Wössmann, L. (2008). Education policy and equality of opportunity. *KYKLOS International Review for Social Sciences*, 61 (2), 279-308.

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

Snijders, T.A.B. (2005). Power and sample size in multilevel modeling. In: B.S. Everitt and D.C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science,* 3, 1570-1573.

Squires, D. (2012). Curriculum alignment research suggests that alignment can improve student achievement. *The Cleaning House*, 85, 129-135.

Su, S.W. (2012). The various concepts of curriculum and the factors involved in curricula-making. *Journal of Language Teaching and Research*, 3(1), 153-158.

Subedi, B. R. (2005*). A demonstration of the three-level hierarchical generalized linear model applied to educational research*. (Unpublished doctoral dissertation). The Florida State University: Tallahassee, FL.

Summers, A. A. & Wolfe, B. L. (1975). Which school resources help learning? Efficiency and equity in Philadelphia public schools. *Business Review*, 4-29.

Tabachnick, B. G. ve Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon, Inc.

Taiwo, A.A. & Tyolo, J.B. (2002). The effect of pre-school education on academic performance in primary school: a case study of grade one pupils in Botswana. *International Journal of Educational Development*, 22, 169-180.

TIMSS & PIRLS International Study Center (2013). *About TIMSS & PIRLS*. Lynch School of Education, Boston College IEA International Association for the Evaluation of Educational Achievement. Retrieved 31 January, 2017, from http://timssandpirls.bc.edu/home/pdf/TP_About.pdf

Tucker-Drop, E. M. (2012). Preschools reduce early academic-achievement gaps: A longitudinal twin approach. *Psychological Science*, 1-10. DOI:10.1177/0956797611426728.

U.S. Department of Education (1986). *What works: Research about teaching and learning.* Washington, DC: Author.

Van de Rijt, B., Godfrey, R., Aubrey, C., Van Luit, J.E.H., Grequiere P., Torbeyns, J, …, Tzouriadou, M. (2003). The development of early numeracy in Europe. *Journal of Early Childhood Research*, 1(2), 155-180.

Van de Werfhorst, H. G. & Mijs, J.J.B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *The Annual Review of Sociology*. DOI: 10.1146/annurev.soc.012809.102538

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), pp.461-481.

Woltman, H., Feldstain, A., Mackay, J. C. & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31, 114-128.

Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Educational Research*, 47(1), 21-41. DOI: 10.1080/00313830308609

Yardbrough, D. B., Shulha, L.M., Hopson, R.K. & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage. Retrieved 7 February, 2017, from http://www.jcsee.org/program-evaluation-standards-statements