THE INFLUENCE OF ITEM FORMAT ON TURKISH EIGHTH GRADERS' PERFORMANCE ON THE TIMSS

NERİMAN KÜÇÜKER

BOĞAZİÇİ UNIVERSITY

2017

THE INFLUENCE OF ITEM FORMAT ON TURKISH EIGHTH GRADERS' PERFORMANCE ON THE TIMSS

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

Neriman Küçüker

Boğaziçi University

2017

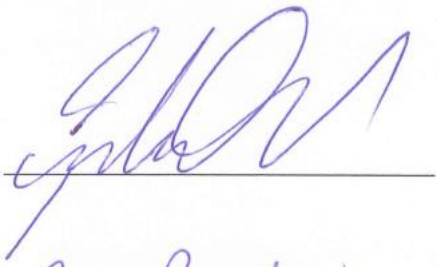
The Influence of Item Format on Turkish Eighth Graders'

Performance on the TIMSS

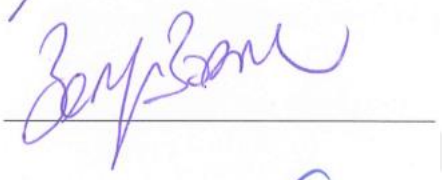
The thesis of Neriman Küçüker

has been approved by:

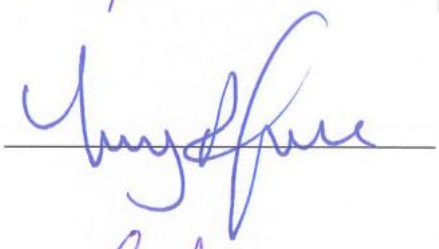
Assoc. Prof. Serkan Özel
(Thesis Advisor)



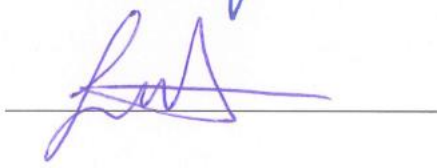
Assist. Prof. Bengü Börkan
(Thesis Co-Advisor)



Prof. Yavuz Akpınar



Assist. Prof. Fatma Aslan-Tutak



Assoc. Prof. M. Sencer Çorlu
(External Member)



May 2017

DECLARATION OF ORIGINALITY

I, Neriman Küçüker, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources 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 25.05.2017

ABSTRACT

The Influence of Item Format on Turkish Eighth Graders'

Performance on the TIMSS

This study examined the influence of item format (multiple choice and constructed response items) on Turkish students' performance on the TIMSS 2011 eighth grade mathematics assessment, together with cognitive (knowing, applying, and reasoning) and content (number, algebra, geometry, and data and chance) domains. In addition, the ranking position changes of Turkey based on different item types were investigated among other participant countries. The study is a descriptive research study. Using the results of the TIMSS 2011 eighth-grade mathematics assessment, students' average percent corrects for each item type were calculated, separately. According to the analyses, Turkish students performed significantly better on multiple-choice items than on constructed response items. Moreover, it was found that the performance of the students was statistically different depending on the cognitive level of the items, and the performances were decreasing from low to high cognitive level of items. It was found that students' performance on data and chance items was significantly higher than their performance on number items as well as on algebra items. However, the interaction effects of item format and cognitive level of items and item format and content domains of items were non-significant. In addition to this, it was observed that, the country rankings based on different item types were significantly similar to the ranking based on overall items, and also the average performance of Turkish students on different item types was generally under international averages, but there was no significant difference between the averages.

ÖZET

8. Sınıf Öğrencilerinin TIMSS Başarı Seviyelerine Madde Tipinin Etkisi

Bu çalışma, Türkiye'deki öğrencilerin TIMSS 2011 matematik testindeki başarılarının madde tipine (çoktan seçmeli ve uzun/kısa cevaplı maddeler) göre nasıl değiştiğini bilişsel (bilgi, uygulama, akıl yürütme) ve içerik (sayılar, cebir, geometri,veri ve olasılık) boyutları ile birlikte incelemiştir. Bunun yanı sıra, madde tipine göre Türkiye'nin diğer katılımcı ülkeler arasındaki başarı sıralamasındaki değişiklik araştırılmıştır. Bu çalışmada betimsel araştırma yöntemi kullanılmıştır. Öğrencilerin TIMSS 2011 8. sınıf matematik başarı testi sonuçları kullanılarak öğrencilerin farklı madde tipindeki soruları doğru cevaplama yüzdelerinin ortalamaları hesaplanmıştır. Yapılan analizler doğrultusunda; öğrencilerin uzun/kısa cevaplı maddelere göre çoktan seçmeli maddelerde belirgin bir şekilde daha başarılı olduğu görülmüştür. Ayrıca, öğrenci başarısının maddelerin bilişsel alan düzeyine göre farklılık gösterdiği ve bilişsel alan düzeyi basitten karmaşığa doğru gittikçe öğrenci başarısının düştüğü saptanmıştır. Öğrencilerin veri ve olasılık konusunda sayılar ve cebir konularına göre daha başarılı olduğu tespit edilmiştir. Ancak bilişsel düzey-madde tipi etkileşimi ile içerik-madde tipi etkileşiminin başarı üzerinde anlamlı bir etkisinin olmadığı görülmüştür. Bunun yanı sıra, madde tipine göre ülke sıralamalarının genel sıralama ile benzerlik gösterdiği, Türkiye'nin başarı ortalamasinin uluslararası ortalamanın genellikle altında olduğu ancak başarı ortalamaları arasında istatistiksel olarak anlamlı bir farklılık olmadığı gözlenmiştir.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Assoc. Prof. Serkan Özel, for his continuous support and guidance during my master's study. I also thank my co-advisor, Assist. Prof. Bengü Börkan, for her advice and support, especially on the statistical analysis of the study. I would like to thank, Prof. Yavuz Akpınar, not only for his valuable comments, but also for his continuous support throughout my master. In addition, I thank the rest of my thesis committee, Prof. Fatma Aslan Tutak, and Assoc. Prof. Sencer Çorlu for their valuable and supportive comments. Also I appreciate Prof. Güzver Yıldıran's encouraging me to start my master's.

Last but not least, many thanks to my family: my mother Bediha Özgür, my sister Leman Aktaş, Emrah Aktaş, Doğa Aktaş, Duru Aktaş, and Zatiye Karakaş for supporting me throughout my life. I will be always in-debted to the members of my family. Specially, I would like to thank my husband, Mehmet Ali Küçüker, for being in my life and for his endless belief and understanding. He encouraged me to find the motivation to finish the study.

Finally, I would like thank all my friends and colleagues for their support and encouragement.

Dedicated to my loves,

my husband

and

my daughter Defne

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
CHAPTER 2: LITERATURE REVIEW 4
2.1 Multiple choice and constructed response items 4
2.2 TIMSS survey 6
2.3 Prior research 9
CHAPTER 3: METHODOLOGY 25
3.1 Participants 25
3.2 Instrument 27
3.3 Research design and procedure 33
3.4 Data analysis 36
CHAPTER 4: RESULTS 38
4.1 Research question 1 38
4.2 Research question 2 40
4.3 Research question 3 45
4.4 Research question 4 52
CHAPTER 5: DISCUSSION 63
5.1 The influence of item format on students' performance 63
5.2 The change in the ranking of Turkey in different item types 66
5.3 Limitations of the study 69
5.4 Recommendations and suggestions for further research 69
APPENDIX A: THE NUMBER OF STUDENTS AND SCHOOLS 71
APPENDIX B: MATHEMATICS CONTENT AND COGNITIVE DOMAINS 72
APPENDIX C: THE NUMBER OF STUDENTS ANSWERED THE ITEM 79
APPENDIX D: ANALYSIS TABLES 83
REFERENCES 87

LIST OF TABLES

Table 1. Turkey Eighth Grade Mathematics Results on TIMSS 1999, 2007 and 2011 8

Table 2. Achievement in Eighth Grade Mathematics Cognitive and Content Domains 9

Table 3. Distribution of the Eighth Grade Students in Turkey 27

Table 4. Cognitive Domains 29

Table 5. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Cognitive Domain 30

Table 6. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Content Domain 30

Table 7. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Item Format, Cognitive Domain and Content Domain 32

Table 8. Countries Participated in the TIMSS 2011 Eighth Grade Mathematics Assessment 35

Table 9. Descriptive Statistics for Students' Performance on MC and CR Items 38

Table 10. Shapiro-Wilk Normality Test Results for the Performances of the Students on MC and CR items 39

Table 11. Descriptive Statistics for the Performances of the Students on Knowing, Applying, and Reasoning Items 41

Table 12. Shapiro-Wilk Normality Test Results for the Performances of the Students on Different Cognitive Level of Items 41

Table 13. One-Way Analysis of Variance for Performances of the Students on Different Cognitive Levels 41

Table 14. Bonferroni Test Results for Comparisons of the Students' Performance on Different Cognitive Levels 42
Table 15. Descriptive Statistics for Students' Performances on MC and CR Items across Cognitive Domains 43
Table 16. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Item Types 44
Table 17. Two-Way Analysis of Variance for Performances of the Students on MC and CR Items on Different Cognitive Levels 45
Table 18. Descriptive Statistics for the Performances of the Students on Number, Algebra, Geometry and Data and Chance Items 46
Table 19. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Content Domains 47
Table 20. Games-Howell Post Hoc Results for Students' Performances on Different Content Items 47
Table 21. Descriptive Statistics for MC Items and CR Items in Different Content Domain 48
Table 22. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Item Types 51
Table 23. Two-Way Analysis of Variance for Performances of the Students on MC and CR Items on Different Content Domains 51
Table 24. The Items were not included to the Analysis of the Country 53
Table 25. Kendall's tau-b Test Results 54
Table 26. Position Changes on Overall Items Compared with MC and CR Items 55
Table 27. Average Percentages Correct at Eighth Grade for 42 Countries across Different Item Type in TIMSS 57

Table 28. Rankings Based on Average Percent Correct for each Item Type 58
Table 29. The Ranking Position Changes of Turkey on Different Item Type Compared with Overall Ranking 60
Table 30. Dunnett's Test Results for International Average and All Countries 61

LIST OF APPENDIX TABLES

Table A1. The Number of Students and Schools Attended TIMSS 2011 Eighth Grade Mathematics Assessment 71
Table A2. Mathematics Content Domains, Topic Areas, and Their Objectives 72
Table A3. Item Labels in each Content and Cognitive Domain 74
Table A4. Items and the Number of Students Answered It 79
Table A5. Average Percentages Correct at Eighth Grade for 42 Countries across Different Item Type in TIMSS 84
Table A6. Rankings Based on Average Percent Correct for Each Item Type 85
Table A7. Dunnett's Test Results for Turkey and All Countries 86

ABBREVIATIONS

CR	Constructed Response
IEA	International Association for the Evaluation of Educational Achievement
LGS	High School Entrance Examination
LYS	Undergraduate Placement Examination
MC	Multiple Choice
MEB	Ministry of National Education
OECD	Organisation for Economic Co-operation and Development
OKS	Middle Schools Selection and Placement Examination
ÖSS	Student Selection Examination
ÖSYM	Student Selection and Placement Center
PISA	Programme for International Student Assessment
SBS	Level Determination Examination
TEOG	Transition From Elementary to Middle School
TIMSS	Trends in Mathematics and Science Study
YGS	Higher Education Examination

CHAPTER 1

INTRODUCTION

Students get evaluated using different examination types such as oral test, formal written test and multiple choice test. Large-scale assessments generally use both multiple choice (MC) and constructed response (CR) questions.

In international tests, Lapointe, Askew, and Mead, (1992) reported that "the testing format –multiple-choice and short answer questions– is not equally familiar to students from all countries" (p.11). In the United States, students perform well in MC questions; however, students are familiar to CR questions in most of the European countries (Wolf, 1998). In Turkey, students take an MC exam before high school and university which are conducted by Ministry of National Education (MEB) and Student Selection and Placement Center (ÖSYM), respectively. In addition, Karakuş (2010), Çakan (2004), and Öztürk, Yalvaç Hastürk, and Demir (2013) reported that teachers use mostly MC test as an assessment method. Hence, students are familiar to MC tests in Turkey. The main question arises at this point: Does student's familiarity with item format have any influence on their performance? On the other side, ÖSYM started to ask CR questions in Undergraduate Placement Examination (Lisans Yerleştirme Sınavı [LYS]) by 2017. Therefore, the students' achievement on CR items becomes more important. Until now, there has been no study in Turkey related with the performance of the students in different item formats. Another question is that: How does Turkish students' performance change in different item formats (MC and CR)?

Cognitive levels of the items have an importance in order to evaluate the learning of the students. Another important aspect of the items is their content

domain. Thus, cognitive and content domain of the items should be taken in consideration for making accurate comparison. Hence, this study focuses on also cognitive and content domains of the items.

In order to view the students' performance in broad perspective, we need to compare Turkish students' performance with other nations. Internatıonal assessments such as Programme for International Student Assessment (PISA) and Trends in Mathematics and Science Study (TIMSS) provide valuable information that helps countries monitor and evaluate their educational systems across time and across grades. TIMSS aims improving teaching and learning in mathematics and science. Therefore, TIMSS assesses students' understanding on mathematics and science at multiple levels. On the other hand, PISA tests the students' skills and knowledge in reading literacy, mathematical literacy and scientific literacy. Hence, TIMSS results become more proper for the research. In view of this aspect, the last question is occurred that how the position of Turkey chances looking at the students' mathematics performance on different item formats in TIMSS 2011.

The main purpose of the study is to determine whether the change in format of mathematics items (MC and CR) within TIMSS 2011 makes difference on students' performance in Turkey. Another aim of the study is whether Turkey's and other countries' (participated in the TIMSS 2011 eighth grade mathematics assessment) position in the rank ordering remains stable when students' mathematics performance on different item sets is compared.

In the light of the purposes stated above, the following research questions were determined: This study will try to search for answers of the following research questions:

1. What is the difference between Turkish students' performances on MC items and CR items of the TIMSS 2011 eighth grade mathematics assessment?
2. What is the difference between Turkish students' performances on MC items and CR items in different cognitive level of the TIMSS 2011 eighth grade mathematics assessment?
3. What is the difference between Turkish students' performances on MC items and CR items in different content level of the TIMSS 2011 eighth grade mathematics assessment?
4. What is the difference between countries rank when looking at students' performances on different item type (item format, cognitive and content domains) of the TIMSS 2011 eighth grade mathematics assessment?

CHAPTER 2

LITERATURE REVIEW

2.1 Multiple choice and constructed response items

The most common multiple choice test consists of one stem and a list of possible answers, one of these choices is correct and the others are incorrect, called distractors. Test takers select the correct answer from the choices (Jennings & Bush, 2006; Kastner & Stangl, 2011). In MC test, examinees can derive the correct answer by a process of elimination; however, well prepared MC questions or selecting more than one correct answer reduce the chances of guessing a question correctly and increase the reliability and validity of the test (Bush, 1999; Foster, 2005; Tekin 2003; Temizkan & Sallabaş, 2011).

MC exams can be graded by using automated scoring therefore the grading process is not only fast and inexpensive but also reliable which means scoring bias do not exist and cross-making is not necessary (Bennett, Ward, Rock & LaHart, 1990; Farthing, Jones & McPhee, 1998; Lissitz & Hou, 2008; Livingston, 2009). A large amount of questions can be asked in MC tests in a limited amount of test time, so large amount of curriculum content are covered by the test and provides a good sample of the test taker's knowledge (Bennett et al., 1990; Livingston, 2009; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2010; Walstad & Becker, 1994). Also the large number of questions increases the reliability and validity of the test (Temizkan & Sallabaş, 2011).

In CR questions, test takers should form their own answers such as a short answer, an essay, a diagram, an explanation of a procedure, the solution of a

mathematical problem (Lukhele, Thissen, & Wainer, 1994). In CR tests, guessing is minimized, so that they are more reliable (Livingston, 2009).

The disadvantages of the CR exams are that the cost of testing is expensive, and the time required to compute and report the scores is more (Lissitz & Hou, 2008; Livingston, 2009; Ventouras et al., 2010). Moreover, it cannot be answered many CR questions in a given time because answering CR questions takes longer time (Livingston, 2009). Hence, CR tests have only few questions (Ventouras et al., 2010) and this reduces the content validity of the test (Tekin, 2003).

MC questions provide little information about students' understanding (Darling-Hammond, 1994). On the other hand, CR questions examine a deeper understanding of the subject (Bacon, 2003).

There are several studies about similarity and dissimilarity of MC and CR tests. Elley and Mangubhai (1992) found that item format (MC and open-ended) had no significant impact on the outcomes of their reading literacy study. They reported that the students, who did well on one test, did well on the other. On the other hand, Temizkan and Sallabaş (2011) found that there was a significant difference in reading comprehension of the students. Students were more successful in MC test than open-ended question test. Also, O'Leary (2001) pointed out that test format will make a difference in performance.

Walstad and Becker (1994) made an analysis using the results of an Advanced Replacement Economics test (macro exam and micro exam) which consists of two sections (50 multiple-choice items and one essay question for macro exam and 50 multiple-choice item and two essay questions for micro exam). In grading of MC exam, they calculated by summing the number of right and subtracting one fourth of the number of false. It was reported that MC test predicted

78% of the micro grade and 72% of the macro grade. In other words, MC and essay exams do not measure the same construct.

Martinez (1999) and Hancock (1994) found that there was no overall equality between MC and CR test for the higher levels of Bloom's Taxonomy (for further information see Bloom, 1956). They stated that MC and CR measure the same level of knowledge only for remembering, understanding, applying and analysing cognitive dimensions.

Livingston (2009) stated about academic subjects that, the students who are stronger in the skills measured by MC items tend to be stronger in the skills measured by CR items but this is not valid for the opposite. Additionally, complex skills and types of knowledge cannot be measured by MC tests (Livingston, 2009).

Hastedt (2004) compared students' scores on MC and CR items based on different item response models (percent correct, RASCH item parameters, and Item Response Theory [IRT] item difficulties). According to percent correct, the results were found 38.3% for CR items and 49.1% for MC items. In other words, CR items were harder than MC items. When RASCH model was applied, it was found that CR items were harder than MC items. When item difficulty was calculated, the values of -0.18 and 0.01 were obtained for CR and MC items, respectively. It was showed that CR items were harder than MC items but the difference was less. It was concluded that, when guessing was not taken to account (RASCH and percent correct) MC items were easier than CR items, and the difference became marginal.

2.2 TIMSS survey

Trend in International Mathematics and Science Study (TIMSS) surveys have been conducted by the International Association for the Evaluation of Educational

Achievement (IEA) on a regular 4-year cycle at fourth and eighth grade (Foy, Arora, & Stanco, 2013).

TIMSS collects data not only mathematics and science achievement but also student, home, teacher, school, and curricular background. Foy et al. (2013a) defined the aim of the survey as "to provide comparative perspectives on trends in mathematics and science achievement in the context of different educational systems, school organizational approaches, and instructional practices" (p.1). In addition, TIMSS enable the countries to evaluate their educational policies and practices in time.

TIMSS gets data by the help of the following questionnaires:

1. Student questionnaires
2. Home questionnaire
3. Teacher questionnaires
4. School questionnaires
5. Curriculum questionnaires
6. Student Achievement Booklet (mathematics and science)

The first TIMSS survey was in 1995 and has continued with subsequent assessments in 1999, 2003, 2007, 2011 and 2015 (Foy et al., 2013a). Turkey participated in TIMSS 1999 and 2007 at eighth grade, TIMSS 2011 and 2015 at fourth and eighth grades. The evaluation of the last survey (TIMSS 2015) has not been completed, therefore TIMSS 2011 results, the last available data, will be used in this study. Table 1 displays the scores and ranks of Turkey in TIMSS 1999, 2007 and 2011 at eighth grade and Figure 1 shows international average and average of Turkey in TIMSS 1999, 2007, and 2011 at eighth grade mathematics assessment (Beaton,

1998, Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009; Mullis, Martin, Foy, & Arora, 2012).

Table 1. Turkey Eighth Grade Mathematics Results on TIMSS 1999, 2007 and 2011

Survey	Score	Rank	Number of total participant countries
TIMSS 1999	429	32	38
TIMSS 2007	432	30	50
TIMSS 2011	452	24	42

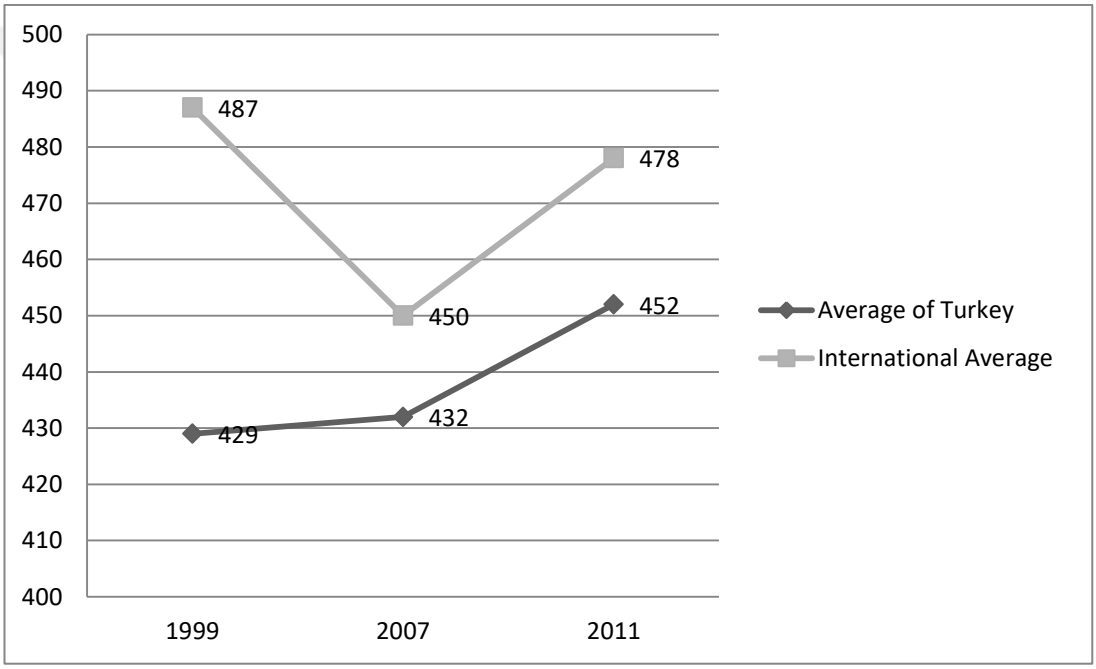


Figure 1 Average of Turkey and international average of TIMSS 1999, 2007 and 2011 eighth grade mathematics scores

Scores of all countries participated in TIMSS 2011 eighth grade mathematics assessment can be seen in Table 2. Table 2 contains overall scores, cognitive domain scores (knowing, applying, and reasoning), content domain scores (number, algebra, geometry, and data and chance) separately.

Table 2. Achievement in Eighth Grade Mathematics Cognitive and Content Domains

Country	Overall avg. scale score	Knowing	Applying	Reasoning	Number	Algebra	Geometry	Data and chance
Korea	613 (2.9)	616 (2.9)	617 (2.9)	612 (2.5)	618 (2.6)	617 (3.2)	612 (2.7)	616 (2.5)
Singapore	611 (3.8)	617 (3.8)	613 (3.9)	604 (4.3)	611 (3.6)	614 (4.1)	609 (3.9)	607 (4.4)
Chinese Tai.	609 (3.2)	611 (3.7)	614 (3.5)	609 (3.4)	598 (3.1)	628 (3.8)	625 (3.7)	584 (3.0)
Hong Kong	586 (3.8)	591 (3.9)	587 (3.7)	580 (3.9)	588 (3.7)	583 (3.9)	597 (4.3)	581 (4.1)
Japan	570 (2.6)	558 (2.7)	574 (2.5)	579 (3.0)	557 (3.0)	570 (3.0)	586 (3.5)	579 (3.0)
Russian Fed.	539 (3.6)	548 (3.6)	538 (3.5)	531 (3.7)	534 (3.2)	556 (3.7)	533 (4.0)	511 (3.9)
Israel	516 (4.1)	516 (4.1)	513 (4.4)	520 (4.0)	518 (4.0)	521 (4.7)	496 (4.6)	515 (4.8)
Finland	514 (2.5)	508 (2.5)	520 (2.5)	512 (2.7)	527 (2.4)	492 (2.9)	502 (2.9)	542 (3.1)
United States	509 (2.6)	519 (2.7)	503 (2.8)	503 (2.7)	514 (3.0)	512 (2.6)	485 (2.7)	527 (3.3)
England	507 (5.5)	501 (5.4)	508 (5.5)	510 (5.5)	512 (5.8)	489 (5.7)	498 (5.7)	543 (6.8)
Hungary	505 (3.5)	507 (3.8)	505 (3.5)	502 (3.7)	510 (3.9)	496 (4.0)	501 (4.1)	517 (4.3)
Australia	505 (5.1)	504 (5.1)	506 (4.8)	506 (4.9)	513 (5.4)	489 (5.3)	499 (5.4)	534 (5.9)
Slovenia	505 (2.2)	508 (2.4)	502 (2.1)	500 (2.7)	511 (2.5)	493 (2.6)	504 (3.1)	518 (3.3)
Lithuania	502 (2.5)	502 (2.6)	508 (2.4)	493 (2.5)	501 (2.5)	492 (2.8)	500 (3.1)	515 (2.8)
Italy	498 (2.4)	494 (2.6)	503 (2.2)	496 (2.6)	496 (2.9)	491 (2.7)	512 (3.5)	499 (3.2)
New Zealand	488 (5.5)	481 (5.6)	491 (5.0)	494 (5.3)	492 (5.9)	472 (5.5)	483 (5.5)	513 (6.7)
Kazakhstan	487 (4.0)	489 (4.4)	484 (4.2)	482 (4.7)	479 (4.0)	506 (4.4)	491 (4.4)	444 (4.5)
Sweden	484 (1.9)	478 (2.0)	489 (2.2)	478 (2.4)	504 (1.8)	459 (2.2)	456 (2.3)	504 (2.7)
Ukraine	479 (3.9)	481 (4.4)	480 (4.3)	467 (4.2)	472 (4.1)	487 (4.4)	476 (4.3)	471 (4.0)
Norway	475 (2.4)	465 (2.5)	480 (2.6)	478 (2.9)	492 (2.8)	432 (2.7)	461 (3.5)	513 (3.6)
Armenia	467 (2.7)	476 (2.9)	458 (3.0)	451 (3.0)	474 (2.4)	496 (2.8)	450 (3.3)	376 (3.7)
Romania	458 (4.0)	460 (4.4)	454 (3.9)	455 (4.0)	448 (4.1)	477 (4.3)	453 (4.5)	429 (4.0)
U. Arab E.	456 (2.1)	467 (2.2)	442 (2.2)	449 (2.1)	459 (2.2)	468 (2.2)	431 (2.4)	440 (2.4)
Turkey	452 (3.9)	441 (4.1)	459 (4.0)	465 (4.7)	435 (3.9)	455 (4.2)	454 (4.3)	467 (4.0)
Lebanon	449 (3.7)	464 (3.9)	436 (4.1)	426 (4.7)	451 (3.8)	471 (3.8)	447 (3.8)	393 (5.2)
Malaysia	440 (5.4)	444 (5.7)	439 (5.2)	426 (5.5)	451 (5.8)	430 (5.2)	432 (6.4)	429 (5.3)
Georgia	431 (3.8)	438 (4.2)	425 (3.6)	414 (4.2)	435 (3.5)	450 (3.8)	406 (4.2)	392 (4.5)
Thailand	427 (4.3)	423 (4.7)	428 (4.1)	429 (4.3)	425 (4.6)	425 (4.3)	415 (5.4)	431 (4.1)
Macedonia	426 (5.2)	430 (5.6)	417 (5.2)	424 (5.9)	418 (5.1)	448 (5.3)	419 (6.0)	389 (5.9)
Tunisia	425 (2.8)	425 (2.8)	421 (2.9)	423 (2.7)	431 (2.8)	419 (2.9)	426 (3.2)	398 (3.3)
Chile	416 (2.6)	405 (2.9)	425 (2.5)	422 (2.8)	413 (2.9)	403 (3.6)	419 (3.1)	426 (3.1)
Iran	415 (4.3)	410 (4.4)	411 (4.6)	428 (4.3)	402 (4.9)	422 (4.3)	437 (4.8)	393 (4.9)
Qatar	410 (3.1)	418 (2.9)	396 (3.3)	406 (3.3)	408 (3.4)	425 (2.8)	387 (3.6)	390 (3.6)
Bahrain	409 (2.0)	411 (2.4)	400 (2.4)	415 (2.1)	397 (1.7)	424 (1.7)	398 (2.6)	407 (2.6)
Jordan	406 (3.7)	405 (4.3)	397 (3.8)	416 (3.8)	390 (3.8)	432 (3.9)	407 (3.7)	379 (3.7)
Palestinian	404 (3.5)	406 (3.5)	397 (3.5)	404 (4.1)	400 (3.4)	419 (3.3)	416 (3.6)	368 (3.6)
Saudi Arabia	394 (4.6)	402 (4.6)	375 (4.8)	388 (4.7)	393 (4.8)	399 (4.9)	364 (5.3)	387 (5.1)
Indonesia	386 (4.3)	378 (4.8)	384 (4.7)	388 (3.8)	375 (4.8)	392 (3.8)	377 (5.3)	376 (4.8)
Syrian A. R.	380 (4.5)	374 (4.4)	379 (4.2)	371 (5.4)	373 (4.0)	391 (4.9)	386 (5.0)	343 (4.7)
Morocco	371 (2.0)	363 (2.2)	378 (1.9)	357 (2.7)	379 (2.6)	357 (2.7)	390 (2.5)	332 (2.0)
Oman	366 (2.8)	365 (3.0)	360 (3.0)	369 (2.8)	351 (3.0)	383 (2.8)	377 (2.7)	342 (3.1)
Ghana	331 (4.3)	331 (4.4)	316 (4.1)	324 (4.8)	321 (4.5)	358 (4.0)	315 (4.3)	296 (4.5)

(Mullis, Martin, Foy, & Arora, 2012)

2.3 Prior research

In Turkey, there are two types of assessments: national and local. Local examinations are applied by teachers such as oral exam, written exam, homework, and project in order to evaluate students. On the other hand, national exams are conducted by ÖSYM and MEB to place students in higher education institutions (for secondary schools or universities) such as Higher Education Examination (Yükseköğretime Geçis Sınavı [YGS]), Undergraduate Placement Examination (Lisans Yerleştirme

Sınavı [LYS]), Student Selection Examination (Öğrenci Seçme Sınavı [ÖSS]), Transition From Elementary to Middle School (Temel Eğitimden Ortaöğretime Geçiş Sistemi [TEOG]), Level Determination Examination (Seviye Belirleme Sınavı [SBS]), Middle Schools Selection and Placement Examination (Ortaöğretime Geçiş Sınavı [OKS]), High School Entrance Examination (Liselere Giriş Sınavı [LGS]).

National assessments provide an estimate of the achievement level of students in the educational system and a curriculum area. Similarly, international assessments have also same purpose, besides giving information about a nation's educational system in relation to one or more other systems.

This section contained prior studies on item format, cognitive, and content domain of the questions in the local, national and international assessments.

2.3.1 Studies on the item format in the local assessment

Karakuş (2010) stated that primary school mathematics teachers mostly used formal written tests and MC tests as classroom assessment. Besides, Çakan (2004) reported that elementary school teachers used mostly MC items, written exams and short answer questions (56%, 49%, and 46%, respectively), while secondary school teachers preferred to use mostly written exams, short answer questions, and MC items (70%, 31%, and 30%, respectively). Also, Öztürk et al. (2013) revealed that all of the primary school teachers at science and technology course in their study used MC items as an assessment method.

2.3.2 Studies on the cognitive domains of the questions

There are many studies related with analyzing the questions in the national exams, in the teachers' exam, on the textbooks, and the objectives of the curriculum in different grade and course in terms of the cognitive levels.

2.3.2.1 Studies related with cognitive levels of Bloom's taxonomy

In this section of the study includes studies analyzed the questions and the objectives of the curriculum in terms of the cognitive levels of Bloom's taxonomy.

Köğce (2005) analyzed mathematics questions in ÖSS in the years 1995-2004 and high school mathematics exam questions at five different high schools; general lycee, technical and multi program high school, vocational and commercial high school, Anatolian high school, and science high school. According to the findings, questions were asked in generally application (47%) and comprehension (32%) levels in ÖSS and there were no questions at knowledge, synthesis, and evaluation levels in each ÖSS exam. Questions at higher cognitive levels increased firstly in 1999 to 20% and then in 2002 to 30%. On the other hand, questions were generally at comprehension level in general lycees, technical and multi program high schools, and vocational and commercial high schools (60%, 72% and 64%, respectively). Besides, questions at application and analysis levels were asked generally in Anatolian high schools and science high schools.

Keleş and Karadeniz (2015) classified mathematics and geometry questions in ÖSS, YGS, and LYS carried out between 2006 and 2012 in terms of Bloom's Revised Taxonomy. According to the study, mathematics and geometry questions were asked mostly at applying and analyzing levels. There was no question at remembering cognitive level. Besides, more than half of the questions were at lower

cognitive levels. The results revealed that, there was not a significant difference between the years according to the dimension of cognitive process of mathematics and geometry questions asked in the university entrance exams.

Keskin and Aydın (2011) investigated cognitive levels of sixth grade SBS biology exam questions in 2008 and 2009. The results showed that most of the questions were at understanding level, and there was no question at analyzing, evaluating and creating levels. Also each exam consisted of only one question at applying level.

Erman (2008) analyzed cognitive level of the OKS history questions in 2003-2006. According to the findings, most of the questions were at comprehension level (72.5%), and the rest 27.5% of the questions were at analysis level.

Koç, Sönmez, and Çiftçi (2013) analyzed cognitive levels of the geography questions at the university entrance exams in the years 2008-2011. According to the results, questions were generally at understanding level in 2008, while at understanding, analysis and evaluating levels in 2009, and also understanding, remembering and analyzing levels in 2010 and 2011. Also, 75-80% of the questions were at lower cognitive levels.

The level of students' thinking depends on the type of questions asked by their teachers. If students expose to questions at lower cognitive level, they will be encouraged to think at a simple level. On the other hand, when questions are asked at higher cognitive levels, students are forced to be more creative and interrogator (Çepni, Ayvacı, & Keleş, 2001). There are many studies related with analyzing teacher' exam questions in different grade and course in terms of the cognitive levels of Bloom's Taxonomy.

Güler, Özdemir, and Dikici (2012) analyzed 6th through 8th grades mathematics exam questions and SBS mathematics questions. According to the findings, all of the sixth grade mathematics exam questions were at lower cognitive levels and questions were asked generally at knowledge level (66.6%). Most of the seventh grade exam questions were at application level (39.1%) and 85.8% of the questions were asked at lower cognitive levels. Questions were generally at comprehension level (55.4%) in eighth grade exams, and 88.7% of the questions were asked at lower cognitive level. On the other hand, most of the questions were asked at application level in SBS exams. In SBS 6th, 7th, and 8th grade exams, 18.75%, 33.2% and 35% of the questions at higher cognitive levels, respectively. The results showed that there was no question at evaluation level among both of the teachers' and SBS exam questions.

Köğce and Baki (2009) studied on high school mathematics exam questions and mathematics questions in ÖSS. They stated that, there was no difference between percentages of science high school mathematics exam questions and ÖSS mathematics questions in terms of cognitive levels. Moreover, Dursun and Aydin-Parim (2014) reported that ninth grade mathematics exam questions and YGS mathematics questions were mostly at application level (73.2% and 84.3%, respectively). Additionally, less than 20% of the mathematics questions were asked higher cognitive levels and, there was no question at synthesis and evaluation levels.

Azar (2005) compared physics questions at the university entrance exams in the years 2000-2003 with ninth and tenth grade physics exam questions at three different high schools; general lycee, Anatolian high school, vocational and commercial high school. According to the study, most of the questions at application and analysis levels (56.33% and 35.53%, respectively). However, high school

physics exam questions at schools were generally at application level (74.64%) and only 5% of the questions were at higher cognitive levels. Also, there was no question at evaluation level in each type of school. Hence it was found that, there was a significant difference between ÖSS physics questions and high school physics exam questions in terms of the cognitive levels of the questions. Moreover, Karaman (2005) studied on physics questions in ÖSS in the years 2001-2002 and high school physics exam questions at four different high schools; general lycee, vocational high school, Anatolian high school, and science high school and the findings were similar with Azar's (2005) study. The findings showed that, teachers asked questions generally at application (62%) and comprehension (18%) levels, and 89% of the questions were at lower cognitive levels.

Sesli (2007) studied on biology questions in ÖSS in the years 1997-2006 and high school biology exam questions at four different high schools; general lycee, vocational high school, Anatolian high school, and science high school. According to the results, biology questions in ÖSS were generally at application, analysis and synthesis cognitive levels, and at least 50% of the questions were at higher cognitive levels. On the other hand, questions were asked generally at knowledge level in vocational high schools, at comprehension level in general lycees, at comprehension and application levels in Anatolian high school, and at application level in science high school. The study showed that there was no difference between percentages of high school biology exam questions and ÖSS biology questions in terms of cognitive levels.

Dindar and Demir (2006) analyzed fifth grade science exam questions. According to their findings, questions related with knowledge were asked mostly (68.63%) and only 1.25% of the questions were at higher cognitive levels.

Ayvacı and Şahin (2009) analyzed both of the science course exam questions and science teachers' questions asked in the class. According to their findings, most of the exam questions and teachers' questions were at knowledge level (47% and 71%, respectively), and also 99% of the questions were at lower cognitive levels. Also it was seen that, teachers in general lycees and vocational high schools asked no question at higher cognitive levels. Moreover, Özcan and Oluk (2007) compared science exam questions. According to the study, questions were asked generally at knowledge (39%) and application (32%) levels. Similar to the results of Ayvacı and Şahin's (2009) study, most of the questions were at lower cognitive levels (96%). Additionally, Gündüz (2009) investigated 6-8th grades science and technology teachers' exam questions. According to the study, the questions were generally at knowledge level (64.65%) and 92.19% of the questions were at lower cognitive levels. It was also stated that the distribution of the questions in terms of the cognitive levels did not depend on the format of the question. In other words, 92% of the both multiple choice questions, and open-ended questions were at lower cognitive levels. Besides, Tanık and Saraçoğlu (2011) analyzed cognitive levels of 6-8th grade science and technology exam questions in terms of Bloom's Revised Taxonomy. The results showed that most of the questions were at lower cognitive levels (90.87%) and questions were generally at remembering level (51.6%) and understanding level (33.1%). However, there was no question at evaluating and creating levels. According to the study, for each grade the results were similar. Furthermore, Gökulu (2015) investigated science and technology questions on both TEOG exam and teachers' exams. The study revealed that, 71% of the science and technology exam questions were at remembering and understanding levels, and 22% of them were at applying level, while only 7% of them were at higher cognitive

levels. On the other hand, 50% of the TEOG science and technology questions were asked at remembering and understanding levels, 23% of them at applying level, and 30% of them at higher cognitive levels.

Özmen and Karamustafaoğlu (2006) carried out tenth grade physics and chemistry courses' exam questions in two different type of school; Anatolian high school and general lycee. According to the results, exam questions were generally at application level for both of the courses and the schools. Although physics and chemistry exam questions were asked generally at lower cognitive levels both in Anatolian high schools (84% and 88%, respectively) and general lycees (99% and 98%, respectively), but no question was asked in general lycees at synthesis and evaluation levels.

Karamustafaoğlu, Sevim, Karamustafaoğlu, and Çepni (2003) investigated high school chemistry exam questions at three different high schools; general lycee, Anatolian high school, vocational and commercial high school. It was found that questions were generally at lower cognitive level (96%). There were no question at synthesis and evaluation levels in both general lycees and vocational and commercial high schools. Also, there was no statistical difference between schools.

Akpınar (2003) analyzed ninth grade geography exam questions. According to the study, questions were asked generally at knowledge and comprehension levels, and also 90% of the questions were at lower cognitive levels.

Baysen (2006) determined the cognitive levels of questions asked by the teachers in the lesson. He revealed that 56% of the questions were at knowledge cognitive level.

According to Seven's (2001) study, in order to reach the aims of the curriculum, textbooks are mostly used tools (72.64%) for each lesson to support

teaching and learning in Turkey. The studies of Beckmann (2004), and Zhu and Fan (2006) showed that when one country's questions in mathematics textbooks are at higher cognitive level, that increases the students' achievement in TIMSS.

Çevik (2010) compared cognitive levels of the questions in 9-11th grades physics textbooks in 2000-2008 and ÖSS physics questions in the years of 2000-2008. The study was found that 13% of the physics textbook questions were at higher cognitive levels; on the other hand, 43% of the ÖSS physics questions were at higher cognitive levels. According to the results, there was a significant difference in terms of cognitive levels of the questions. In each grade textbook, questions were asked mostly at application level (approximately 50%). However, ÖSS physics questions were asked generally at comprehension (28.5%), application (27%) and evaluation levels (24.5%). Also there was any question at synthesis level both in ÖSS and textbook questions.

Küçük (2008) analyzed sixth grade Turkish workbook. The activities part of the book consisted of mostly at comprehension (36% of the questions) and application levels (23.2% of the questions). On the other hand, questions were generally at comprehension level (57%) and knowledge level (26%) in the end of unit tests.

The following studies analyzed the objectives of the curriculum in term of Bloom's Revised Taxonomy. Kablan, Baran, and Hazer (2013) investigated the objectives in the mathematics curriculum for 6-8th grades. The findings of the study indicated that, there was no objective at remembering cognitive process, while the most of the objectives were at understanding and applying levels (49% and 38%, respectively). The study compared the distribution of the objectives on cognitive process by grades. According to the results, objectives on the sixth and eighth grade

mathematics curriculum were mostly at understanding level (57% and 58%, respectively), while on the seventh grade objectives were mostly at applying level (52%). However, there was no significant difference between the objectives at higher cognitive levels by grade. The study also analyzed the cognitive levels of the objectives across content. According to the findings, there was no significant difference between contents in terms of the cognitive levels of the objectives. Geometry objectives were mostly at understanding level (61%), while algebra objective were at applying (57%). Besides, data and chance objectives had more objectives than other contents at higher cognitive levels (23%).

Bekdemir and Selim (2008) classified the objectives of algebra in 6-8th grade mathematics curriculum. According to the study, the objectives were represented generally with the conceptual knowledge and procedural knowledge dimensions in the knowledge dimension. It was reported that there was no objective in meta-cognitive knowledge dimension. The objectives were generally at understanding and applying cognitive levels.

Arı and Gökler (2012) compared the objectives of 8th grade science and technology curriculum and SBS 8th grade science and technology exam questions. The results showed that, the objectives were represented with conceptual knowledge dimension (82%) in the knowledge dimension. It was found that there was any objective in meta-cognitive knowledge dimension. Most of the objectives were at lower cognitive levels (68.4%). Similarly, questions were asked generally at conceptual knowledge dimension (80.7%) in the knowledge dimension, and there was any question in meta-cognitive knowledge dimension. Also, most of the questions were at lower cognitive levels (57.5%).

Gezer, Şahin, Sünkür, and Meral (2014) analyzed the objectives of eighth grade History of Turkish Revolution and Kemalism. According to the results, 92% of the objectives were formed by the conceptual knowledge dimension. It was found that there was no objective in procedural knowledge and meta-cognitive knowledge dimensions. Most of the objectives were at higher cognitive levels (66.7%); however, the objectives in the curriculum were not represented in a balanced way.

Eroğlu and Kuzu (2014) investigated and classified grammar questions in the Turkish workbook for 6-8th grades and objectives in the Teacher's guide book. According to the study, there was no objective and question at higher cognitive levels. Grammar questions and objectives were generally at applying level (49% and 53.42%, respectively) and understanding level (40.98% and 45.30%, respectively).

2.3.2.2 Studies related with cognitive levels of other taxonomies

Delil and Tetik (2015) classified and compared mathematics questions at LGS, SBS, OKS, and TEOG examinations in the years 1998-2015 in terms of TIMSS-2015 cognitive domains. The study findings revealed that most of the questions were asked at applying cognitive domain (58% of the questions), whereas 13% of the questions were at reasoning cognitive domain. The study compared the results before and after the reform in 2005. The results stated that, while the number of knowing and reasoning questions were decreasing, the number of applying questions increased. However, there was no significant difference between the distributions of the questions in terms of cognitive domains year by year.

Güner (2015) compared the questions on geometry and data and chance content domains in the 6th through 8th grade mathematics textbooks published before and after the reform using TIMSS 2007 cognitive domains. According to this

study, 10% of geometry and 20% of data and chance questions were classified at reasoning cognitive level. The results of the study revealed that, there was a significant difference between distribution of the knowing and applying questions in the textbooks published in 2004 and 2008 on geometry. The number of geometry questions at knowing level decreased by 2008, while the number of applying geometry questions increased. After the reform, the percentages of the geometry questions at knowing, applying and reasoning levels were 40%, 50% and 10%, respectively. On the other hand, there was no significant difference between distribution of the data and chance questions in terms of cognitive domains. In 2008, the percentages of the data and chance questions at knowing, applying and reasoning levels were 33%, 46% and 21%, respectively.

Coşar (2010) studied on cognitive level of the questions in sixth grade mathematics textbooks in terms TIMSS 2007 cognitive domains. It was reported that, 71.25% of the questions were at knowing level, 25.05% of them were at applying level and the rest 3.7% of the at reasoning level. Also the distribution of the questions was unbalanced in each learning unit in terms of cognitive levels.

Aydoğdu Iskenderoğlu and Baki (2011) examined eighth grade mathematics textbook questions in terms of Program for International Student Assessment (PISA) mathematics competency scale. According to the results, there were questions, problems, exercises and examples at Level 1, 2, 3, and 4, and they were mostly at Level 2 (47%).

Delil (2006) classified the geometry problems in 6-8th grade mathematics textbooks in terms of TIMSS 2003 cognitive domains. It was found that, computing and applying (72%) were most frequent behaviors that the problems in the textbooks required.

2.3.3 Studies on the content domains of the questions

Incikabi (2012) compared SBS with TIMSS eighth grade assessments in terms of distribution of content domains, cognitive domains and item types (MC and CR). The results revealed that there was a considerable overlap in the content, differences existed in the distribution of items with respect to cognitive domains and item types. All types of questions in SBS were MC, but in the TIMSS assessment half of the questions were MC. The distribution of the questions in SBS in terms of content domains (number, algebra, geometry and data and chance) were 20%, 35%, 35%, and 10%, whereas in TIMSS 30%, 30%, 20%, and 20% , respectively. In other word, most of the items in SBS were related with geometry and algebra content domains. Besides, the percentages of the questions in terms of cognitive domains (knowing, applying, and reasoning) were 33, 55 and 12 in SBS; whereas 35, 40 and 25 in TIMSS, respectively. In SBS assessment questions were dominated highly by knowing and applying questions.

2.3.4 Studies on performance of the students

Dursun and Aydin-Parim (2014) stated that students answered correctly 35.5% of the questions at knowing level, 38.2% of the questions at application level and 41.35% of the questions at analysis level in YGS mathematics.

Özmen and Karamustafaoğlu (2006) investigated the performance of the tenth grade students on the subject energy from both Anatolian high schools and general lycees. According to the findings, students form general lycees answered correctly 74% of the questions at knowledge level, 72.3 % of the questions at comprehension level, 62.2% of the questions at application level, and 5-21% of the questions at the higher cognitive levels. Moreover, the students from Anatolian high

schools answered correctly 79.5% of the questions at knowledge level, 79% of the questions at comprehension level, 67% of the questions at application level, and 11-19% of the questions at higher cognitive levels. The study showed that, there was not a significant difference between the performances of the students in terms of type of school.

Ceylan (2013) examined science content and cognitive domain scores with regard to low-, medium- and high performing schools in Turkey. According to the results, the earth science content domain scores were relatively lower than the other domains in both fourth and eighth grades for three categories of schools. Fourth grade students from low-, medium-, and high performed schools answered correctly 28%, 45%, and 62% of the science items, respectively, and also 27%, 43%, and 63% of the science items correctly were answered by eighth grade students in low-, medium-, and high performed schools, respectively. Moreover, the findings revealed that, fourth grade students' reasoning scores were higher than students' knowing and applying scores, while eighth grade students' knowing score were higher than their applying and reasoning scores. Besides, the study reported that even the students in high performing schools had difficulties to answer open-ended questions and reasoning items correctly at both fourth and eighth grades.

Kaleli-Yılmaz and Hancı (2016) reported that eighth grade students attended the research in Bayburt had generally more success in knowing and applying cognitive levels. However, the percentage of the reasoning items success was low.

Doğan and Tatsuoka (2008) investigated Turkish students' profile on mathematical skills on TIMSS-R. The findings showed that, Turkish students were weak in algebra and probability/statistics. Also the study stated that, students did not

perform well applying rules in algebra, approximation/estimation, solving open-ended problems, recognizing patterns and relationships, and quantitative reading.

2.3.5 Studies on international assessments

Acar (2012) studied on determining the status of Turkey among Organisation for Economic Co-operation and Development (OECD) members and candidate countries according to PISA 2009 results. According to the analyses, there were four clusters and Turkey was clustered along with Bulgaria, Chile, Colombia, Israel, Jordon, Mexico, Dubai (United Arab Emirates), Romania, Serbia, Thailand, Trinidad, and Tobago. The results showed that, average scores of reading competency, mathematics, and science variables differed significantly among clusters.

O´Leary (2002) studied on country rankings across item formats. The study findings revealed that, the ranking of countries for MC items reflected the overall rankings fairly closely. Also, the rankings were more stable across the MC and short answer items than across the MC and extended-response items.

Hastedt and Sibberns (2005) analyzed TIMSS 1995 and 1999 mathematics and science results across item format, and compared rankings based on MC and CR items. The results of TIMSS 1995 science assessment showed that no country performed well on one scale but poorly on another scale, although there were some changes in the rankings. The findings of the study revealed that, all eastern European countries participated TIMSS 1995 performed better on MC items than CR items. Moreover, the study reported that the ranking based on MC items was similarly with international ranking both in TIMSS 1995 and 1999 science assessments, and also both TIMSS 1995 and TIMSS 1999 mathematics rankings based on MC items, CR items and total scale were very similar.

Nixon, and Barth (2014) compared students' performance from five countries on cognitive domains by using some TIMSS items came from TIMSS 1995, 2003, and 2007. According to the study, there was no significantly difference in students' performance in the sample on knowing and applying items except students in Hong Kong. Moreover, the results revealed that changes in the rankings did not show difference in students' performance when items grouped by domain.

Beaton (1998) compared fourth grade students' performance from 24 countries by TIMSS mathematics content sub-areas. The study showed that there is a strong relationship among performances in the different sub-areas. If the students were successful in one sub-area, they were more likely to do well in other sub-areas. In other words, it was reported that the top-scoring countries are success in all sub-areas, the middle-scorers do middling well, and low-scorers perform less well.

Adams, Berezner, and Jakubowski (2010) studied on an analysis of PISA 2006 preferred items ranking. The findings of the study showed that the ranking positions of the several countries could be altered systematically but the sampling error could explain the shifts in ranking positions of these countries. On the other hand, systematic shifts were not observed for most of the countries.

This study extended these researches in three aspects. It included all items and all countries participated in TIMSS 2011 eighth grade mathematics assessment. Also, it examined not only the performance of Turkish students, but also the ranking chances based on item format, cognitive and content domains of items.

CHAPTER 3

METHODOLOGY

The methodology section of this study includes the characteristics of the sample, the instrument, the design of the study, the research hypotheses, variables and operational definitions, the research procedures, and the data analysis.

3.1 Participants

The target population of TIMSS 2011 is students in fourth and eighth years of schooling which is fourth and eighth grade students in most of the participating countries (Joncas, 2013). Since age of starting school varies country to country, some of students are sixth and ninth grade in the target population. TIMSS 2011 included in 63 countries and 14 benchmarking participants. Fifty countries and seven benchmarking participants in fourth grade, 42 countries and 14 benchmarking participants in eighth grade, two countries in sixth grade and three countries in ninth grade attended TIMSS 2011(Joncas, 2013).

TIMSS uses a stratified two-stage cluster sample design (Martin & Mullis, 2012). For the first sampling stage, schools are sampled with probabilities proportional to their size and for the second sampling stage, one or more classes are selected randomly from the target grade of each participating school (Joncas & Foy, 2013). TIMSS are designed to provide valid and reliable measurement of trends in student achievement in countries, while keeping to a minimum the burden on schools, teachers, and students. Each country participating in TIMSS needs a plan for defining its national target population and applying the TIMSS sampling methods to achieve a nationally representative sample of schools and students, and conducts it

contact with the country's National Research Coordinator and the TIMSS sampling experts (Joncas & Foy, 2013). Joncas and Foy report the following:

> To meet the TIMSS & PIRLS standards for sampling precision, national student samples should provide for a standard error no greater than .035 standard deviation units for the country's mean achievement. With a standard deviation of 100 on the TIMSS and PIRLS achievement scales, this standard error corresponds to a 95% confidence interval of ± 7 score points for the achievement mean and of ± 10 score points for the difference between achievement means from successive cycles (e.g., the difference between a country's achievement mean on TIMSS 2007 and TIMSS 2011). Sample estimates of any student-level percentage estimate (e.g., a student background characteristic) should have a confidence intervals of ± 3.5%. (pp. 7)

TIMSS required that all student sample sizes should not be less than 4,000 students. For most countries, the TIMSS precision requirements were met with a school sample of 150 schools and a student sample of 4,000 students for each target grade. Depending on the average class size in the country, one or two classes from each sampled school were chosen (Joncas & Foy, 2013).

The subject of this study is eighth grade students who attended TIMSS 2011 in Turkey. There are 6,928 students (0.6% of the population) and 239 schools in the sample (Joncas, 2013). As it is presented in Table 3, 49.2% of the students are girls and 50.8% of them are boys in the sample which is almost the same proportion with the population of Turkey (Büyüköztürk, Çakan, Tan & Atar, 2014). Also, TIMSS 2011 eighth grade sample sizes in each 42 countries are available in Table A1 (See Appendix A). At the time of testing, the average age is 14 for the students in the sample (Joncas, 2013).

Table 3. Distribution of the Eighth Grade Students in Turkey

	Population		Sample	
Gender	n	%	n	%
Girls	576,497	49.5	3,414	49.2
Boys	589,225	50.5	3,514	50.8
Total	1,165,722	100.0	6,928	100.0

3.2 Instrument

The subject area of the study is eighth grade mathematics, and TIMSS 2011 eighth grade mathematics assessment results were used for this study. In the TIMSS mathematics assessment, a matrix-sampling design is used such that each student is administered only a subset of the entire TIMSS mathematics item pool. By the help of matrix-sampling design in large populations, more efficient estimates can be obtained (Foy, Brossman, & Galia, 2013). The TIMSS mathematics assessment consists of 14 mathematics achievement blocks at eighth grade. Six of them consist of newly developed items (released items) and eight of them include the previous TIMSS mathematics assessment items (secured items). The number of items in each block is between 12 and 18. Two groups of blocks form 14 booklets in which there are 26 to 33 items. Each item appears in two booklets, providing a mechanism for linking together the student responses from the various booklets. Students were allowed 45 minutes in order to complete the TIMSS eighth grade mathematics achievement booklet (Mullis et al., 2009).

Two item formats are used in TIMSS: multiple choice (MC) items and constructed response (CR) items. In TIMSS, MC items consist of four response options and only one of them is correct, while for CR items, students are required to construct a written response, rather than select a response from a set of options (Mullis et al., 2009).

The TIMSS 2011 eighth grade mathematics achievement booklets include totally 217 mathematics items: 118 of them are MC items, 99 of them are CR items which comprise 54% and 46% of the total items, respectively (Mullis et al., 2012). The TIMSS mathematics assessment includes three different cognitive domains (knowing, applying, and reasoning) and four different content domains (number, algebra, geometry, data and chance) at eighth grade.

The first cognitive domain is knowing which involves the facts, concepts, and procedures students need to know. The second one is applying that focuses on the ability of students to apply knowledge and conceptual understanding to solve problems or answer questions. The last one is reasoning which covers the solution of routine problems to contain unfamiliar situations, complex contexts, and multistep problems (Mullis et al., 2009). Table 4 presents cognitive processes required in each cognitive domain.

Table 4. Cognitive Domains

Cognitive Domain	Cognitive Process	
Knowing	Recall	Recall definitions; terminology; number properties; geometric properties; and notation (e.g., a+a+a=3a, a×b=ab).
	Recognize	Recognize mathematical objects, e.g., shapes, numbers, expressions, and quantities. Recognize mathematical entities that are mathematically equivalent (e.g., equivalent familiar fractions, decimals and percents; different orientations of simple geometric figures).
	Compute	Carry out algorithmic procedures for +, −,×,÷, or a combination of these with whole numbers, fractions, decimals and integers. Approximate numbers to estimate computations. Carry out routine algebraic procedures.
	Retrieve	Retrieve information from graphs, tables, or other sources; read simple scales.
	Measure	Use measuring instruments; choose appropriate units of measurement.
	Classify/order	Classify/group objects, shapes, numbers, and expressions according to common properties; make correct decisions about class membership; and order numbers and objects by attributes.
Applying	Select	Select an efficient/appropriate operation, method, or strategy for solving problems where there is a known procedure, algorithm, or method of solution.
	Represent	Display mathematical information and data in diagrams, tables, charts, or graphs, and generate equivalent representations for a given mathematical entity or relationship.
	Model	Generate an appropriate model, such as an equation, geometric figure, or diagram for solving a routine problem.
	Implement	Implement a set of mathematical instructions (e.g., draw shapes and diagrams to given specifications).
	Solve Routine Problems	Solve standard problems similar to those encountered in class. The problems can be in familiar contexts or purely mathematical.
Reasoning	Analyze	Determine, describe, or use relationships between variables or objects in mathematical situations, and make valid inferences from given information.
	Generalize/Specialize	Extend the domain to which the result of mathematical thinking and problem solving is applicable by restating results in more general and more widely applicable terms.
	Integrate/Synthesize	Make connections between different elements of knowledge and related representations, and make linkages between related mathematical ideas. Combine mathematical facts, concepts, and procedures to establish results, and combine results to produce a further result.
	Justify	Provide a justification by reference to known mathematical results or properties.
	Solve Non-routine Problems	Solve problems set in mathematical or real life contexts where students are unlikely to have encountered closely similar items, and apply mathematical facts, concepts, and procedures in unfamiliar or complex contexts.

(Mullis et al., 2009)

Cognitive domains knowledge, applying, and reasoning form 37%, 39%, and 24% of the items in the TIMSS 2011 eighth grade mathematics assessment,

respectively. Table 5 presents the distribution of the TIMSS 2011 mathematics assessment items by cognitive domain (Mullis et al., 2012, Mullis et al., 2009).

Table 5. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Cognitive Domain

Cognitive Domain	Number of MC Items	Number of CR Items	Total		Target percentage[1] (%)
			Number of Items	Percentage (%)	
Knowing	53	27	80	37	35
Applying	47	38	85	39	40
Reasoning	18	34	52	24	25
Total	118	99	217	100	100

[1] Target percentages of testing time devoted to each cognitive domain for the TIMSS 2011 eighth grade assessments.

The distribution of the TIMSS 2011 eighth grade mathematics assessment items by content domain is displayed in Table 6. Number content domain forms 28% of the total items, 32% of the items consisted of algebra, 20% of them are concerned about geometry, and the rest 20% of them are comprised of data and chance (Mullis et al., 2012, Mullis et al., 2009).

Table 6. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Content Domain

Content Domain	Number of MC Items	Number of CR Items	Total		Target percentage[1] (%)
			Number of Items	Percentage (%)	
Number	31	30	61	28	30
Algebra	37	33	70	32	30
Geometry	25	18	43	20	20
Data and Chance	25	18	43	20	20
Total	118	99	217	100	100

[1] Target percentages of testing time devoted to each content domain for the TIMSS 2011 eighth grade assessments.

Each content domain has several topic areas; each topic area is presented as a list of objectives. The topics of number were defined as whole numbers, fractions

and decimals, integers, and ratio, proportion and percent; the topics of algebra were patterns, algebraic expressions, and equations/formulas and functions; the topics of geometry were geometric shapes, geometric measurements, and location and movement; and the topics of data and chance were data organization and representation, data interpretation, and chance (Mullis et al., 2009). The objectives of each topic area can be seen in Table A2 (See Appendix B). Also, item labels in each content and cognitive domain are available in Table A3 (See Appendix B).

More specifically, TIMSS eighth grade mathematics assessment consists of mostly algebra knowing items with 15% of the total items, and geometry knowing items are least with 3%. For each content domain proportion of applying items are same. The percentages of the items in each content and cognitive domain can be seen in Figure 2 (Mullis et al., 2012).

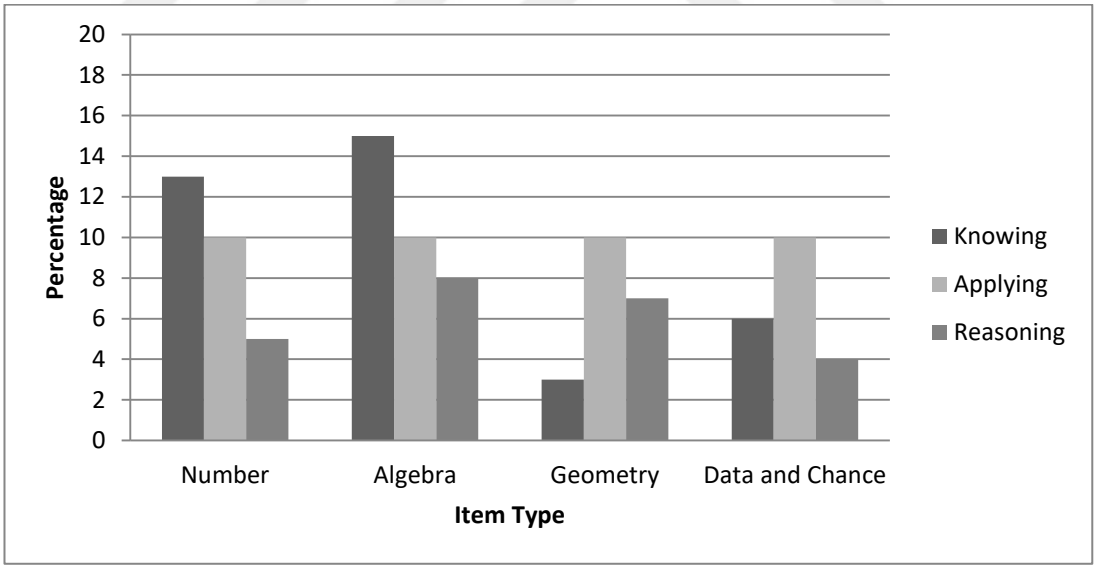


Figure 2 Percentage of the items in each content and cognitive domain in TIMSS 2011 eighth grade mathematics assessment

The distribution of the TIMSS 2011 eighth grade mathematics assessment items by item format, cognitive and content domains is shown in Table 7. In

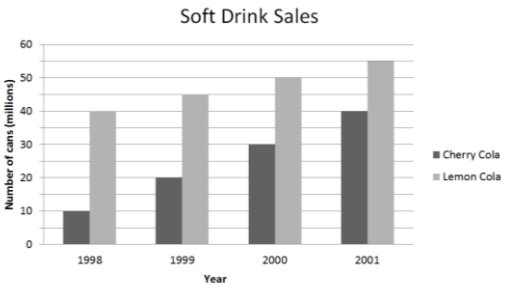
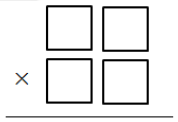
addition, the list of items and the number of students answered each item are available in Table A4 (See Appendix C).

Table 7. Distribution of the TIMSS 2011 Eighth Grade Mathematics Assessment Items by Item Format, Cognitive Domain and Content Domain

Content Domain	Cognitive Domain	Item Format		Total
		Number of MC Items	Number of CR Items	
Number	Knowing	18	11	29
	Applying	10	12	22
	Reasoning	3	7	10
Algebra	Knowing	21	11	33
	Applying	14	7	21
	Reasoning	2	15	17
Geometry	Knowing	5	1	6
	Applying	10	11	21
	Reasoning	10	6	16
Data and Chance	Knowing	9	4	13
	Applying	13	8	21
	Reasoning	3	6	9
Total		118	99	217

(IEA, 2013b)

Examples of item types from each cognitive domain and format in different content levels are provided in Figure 3. Released items of TIMSS eighth grade mathematics assessment can be seen in the webpage (IEA, 2013a).

	Multiple Choice Items	Constructed Response Items
Knowing	M052216 Number (Fractions and Decimals) Which number is equal to $\frac{3}{5}$? A) 0.8 B) 0.6 C) 0.53 D) 0.35	M042198A Algebra (Patterns) $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}$ A. What is the next term in this pattern? Answer: _____
Applying	M052084 Geometry (Geometric Measures) The perimeter of a square is 36 cm. What is the area of this square? A) 81 cm^2 B) 36 cm^2 C) 24 cm^2 D) 18 cm^2	M042169C Data and Chance (Data Interpretation) The Real Burger Company owns 5 restaurants. The number of staff members employed in their 5 restaurants are: 12, 18, 19, 21, and 30 people. C. If the restaurant with 30 staff members increased its number of staff members to 50, how would this affect the median and the mean?
Reasoning	M032721 Data and Chance (Data Interpretation)  The graph shows the sales of two types of soft drink over 4 years. If the sales trends continue for the next 10 years, determine the year in which the sales of Cherry Cola will be the same as the sales of Lemon Cola. A) 2003 B) 2004 C) 2005 D) 2006	M042002 Number (Whole Numbers) Place the four digits 3, 5, 7, and 9 into the boxes below in the positions that would give the greatest result when the two numbers are multiplied. 

(IEA, 2013a)

Figure 3 Examples of TIMSS 2011 eighth grade mathematics items

3.3 Research design and procedure

The study is a descriptive research study. The two independent variables are the countries and the item type. The dependent variable is the performance of the students in the TIMSS 2011 eighth grade mathematics assessment.

3.3.1 Hypotheses of the study

The hypotheses of this study are given as follows:

1. There is a significant difference between the students' performance of MC items and CR items in the TIMSS 2011 eighth grade mathematics assessment.
2. a. There is a significant difference between students' performance on MC and CR items in the TIMSS 2011 eighth grade mathematics assessment.

   b. There is a significant difference between students' performance in different cognitive domain of items in the TIMSS 2011 eighth grade mathematics assessment.

   c. There is a significant interaction effect of item type by item format and its cognitive domain on students' performance in the TIMSS 2011 eighth grade mathematics assessment.
3. a. There is a significant difference between students' performance on MC and CR items in the TIMSS 2011 eighth grade mathematics assessment.

   b. There is a significant difference between students' performance in different content domain of items in the TIMSS 2011 eighth grade mathematics assessment.

   c. There is a significant interaction effect of item type by item format and its content domain on students' performance in the TIMSS 2011 eighth grade mathematics assessment.
4. There is a significant difference in ranking of Turkey compared to other countries who attended the TIMSS 2011 eighth grade mathematics assessment when different item formats are used.

3.3.2 Variables and operational definitions

In this section the operational definition of each variable is given.

Forty two countries participated in the TIMSS 2011 eighth grade mathematics assessment, which is shown in Table 8.

Table 8. Countries Participated in the TIMSS 2011 Eighth Grade Mathematics Assessment

• Armenia	• Indonesia	• Malaysia	• Slovenia
• Australia	• Islamic Republic of Iran	• Morocco	• Sweden
• Bahrain	• Israel	• New Zealand	• Syrian Arab Republic
• Chile	• Italy	• Norway	• Thailand
• Chinese Taipei	• Japan	• Oman	• Tunisia
• England	• Jordan	• Palestinian National Authority	• Turkey
• Finland	• Kazakhstan	• Qatar	• Ukraine
• Georgia	• Republic of Korea	• Romania	• United Arab Emirates
• Ghana	• Lebanon	• Russian Federation	• United States
• Hong Kong SAR	• Lithuania	• Saudi Arabia	
• Hungary	• Republic of Macedonia	• Singapore	

Item type includes three different properties such as its format and its cognitive and content domains.

Item format is a discrete variable which is divided into two categories:

1. Multiple Choice Item
2. Constructed Response Item

Cognitive domain is a discrete variable which is divided into three categories:

1. Knowing
2. Applying
3. Reasoning

Content domain is a discrete variable which is divided into four categories:

1. Number
2. Algebra
3. Geometry

4. Data and Chance

Performance of the students is measured by average percent correct of each item of the TIMSS 2011 eighth grade mathematics assessment. Average percent correct is the arithmetic average of the percentage of students that answered the item correctly, and it was obtained from the TIMSS 2011 International database webpage (IEA, 2013b).

3.3.3 Procedure

The TIMSS & PIRLS International Study Center computed item statistics for all items in the 2011 assessments (Foy, Martin, Mullis, & Stanco, 2013c). The difficulty level of all items (the percentage of students that answered the item correctly) and information about their type (their format, their cognitive and content domains) were obtained from the TIMSS 2011 International Database webpage (IEA, 2013b).

3.4 Data analysis

The principal focus of the analysis is undertaken on the mathematics performance of Turkish eighth grade students in the TIMSS 2011 across different item types. To examine the influence of item type on students' performance, average percent correct was calculated for each item format, cognitive and content domains separately. In addition, average percent correct of MC items in each cognitive and content domain (MC knowing, MC applying, MC reasoning, MC number, MC algebra, MC geometry, MC data and chance items) and CR items in each cognitive and content domain (CR knowing, CR applying, CR reasoning, CR number, CR algebra, CR geometry, CR data and chance items) were calculated separately.

The data obtained from the TIMSS 2011 was analyzed using the independent *t*-test, one-way ANOVA, Bonferroni post-hoc test, Welch test, Games Howell post-hoc test, and two-way ANOVA of the SPSS package program and the level of significance (*p*) is specified as 0.05.

Independent *t*-test was conducted in order to compare students' performance on different item formats. According to Nixon and Barth (2014), independent *t*-test is reasonable, because items are not taken by the same students in the TIMSS and the scores can be treated as independent of each other. One-way ANOVA and Bonferroni post-hoc test were applied to compare students' performance on different cognitive level of item while Welch and Games Howell tests were conducted for examining the performance of students on different content domain of items. Two-way ANOVA was used to compare students' performance on MC and CR items in each cognitive and content domain separately.

Countries can be ranked according to average percent correct which provides a close approximation to the ranking depending on TIMSS scaling methodology (Adams et al., 2010). Finally, 42 countries were ranked according to students' performance on all of the different item types. To identify whether Turkey's ranking position changes are significant, post-hoc Dunnett's test was applied.

CHAPTER 4

RESULTS

The following sections present the findings of the descriptive and inferential statistical analyses under each research question.

4.1 Research question 1

What is the difference between Turkish students' performances on MC items and CR items of the TIMSS 2011 eighth grade mathematics assessment?

H_0: There is no significant difference between the students' performance of MC items and CR items in the TIMSS 2011 eighth grade mathematics assessment.

The descriptive statistics results of Turkish students' performance on MC and CR items of TIMSS eighth grade mathematics assessment are presented in Table 9. Average of the percentage of the students that answered correctly MC items is 45.62% whereas 28.68% for CR items.

Table 9. Descriptive Statistics for Students' Performance on MC and CR Items

Item Type	Number of Items	*M*(%)	SD	Min.	Max.
MC	118	45.62	13.00	14.80	78.60
CR	99	28.68	16.22	0.70	77.30

Independent *t*-test was conducted in order to compare students' performance on MC items with CR items. The test for normality, examining standardized skewness and the Shapiro-Wilks test indicated that the data were normally distributed for MC items but not CR items (See Table 10). Histograms of distributions of students' performance on MC and CR items can be seen in Figure 4 and Figure 5, respectively. Since samples are relatively large, this validation of the

normality assumption compensated (Gravetter & Wallnau, 2012). However, the *Levene's* test revealed that the homogeneity of variance assumption was not observed ($p < .05$).

Table 10. Shapiro-Wilk Normality Test Results for the Performances of the Students on MC and CR items

Groups	Statistic	*df*	*p*
MC	0.989	118	0.488
CR	0.967	99	0.014

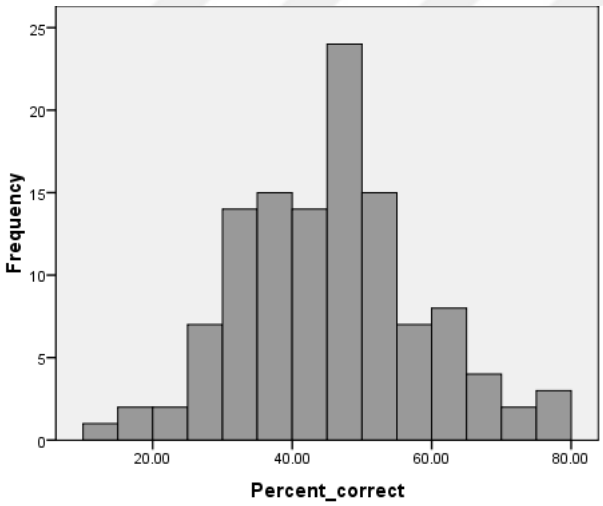


Figure 4 Histogram of the distribution of students' performance on MC items

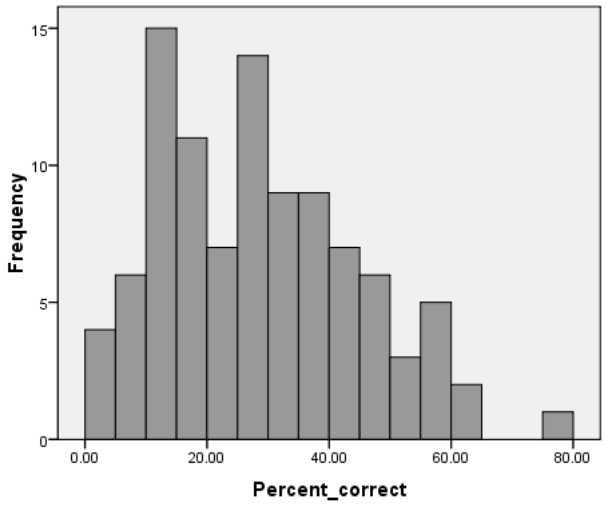


Figure 5 Histogram of the distribution of students' performance on CR items

According to the independent *t*-test results, there is a significant difference between the students' performance on MC items (M = 45.62, SD = 13) and CR items (M = 28.68, SD = 16.22) conditions; $t(186.69) = 8.38$, $p = .00$, Cohen's $d = 1.15$, and as a consequence, H_0 was rejected. These results suggest that the students performed better on MC items than CR items. Specifically, a greater percentage of the students answered mathematics MC items correctly than CR items.

4.2 Research question 2

What is the difference between Turkish students' performances on MC items and CR items in different cognitive level of the TIMSS 2011 eighth grade mathematics assessment?

H_{01} was rejected in the part of research question 1.

H_{02}: There is no significant difference between students' performance in different cognitive domain of items in the TIMSS 2011 eighth grade mathematics assessment.

The performances of the students on knowing, applying, and reasoning items were analyzed separately. It was seen that the percentage of the students' correctly answering knowing items is the most with 44.02%. This is followed by performances of applying and reasoning items with 37.55% and 29.01%, respectively. The descriptive statistics results of students' performance on knowing, applying and reasoning items of the TIMSS eighth grade mathematics assessment can be seen in Table 11.

Table 11. Descriptive Statistics for the Performances of the Students on Knowing, Applying, and Reasoning Items

Item Type	Number of Items	*M* (%)	*SD*	min.	max.
Knowing	80	44.02	16.53	3.30	78.60
Applying	85	37.55	15.32	1.70	70.90
Reasoning	52	29.01	15.72	0.70	63.20

Prior to conducting a one-way ANOVA test, the normality of the data sets and the homogeneity of variances were assessed. A Shapiro-Wilk's test ($p > .05$) and a visual inspection of their histograms, normal Q-Q plots and box plots showed that the students' performances were approximately normally distributed for each group of sample (See Table 12). A *Levene's* test verified the equality of variances in the samples ($p > .05$). The one-way ANOVA test [$F(2,214) = 14.12$, $p < .01$] revealed that the performance of the students were statistically different for the different cognitive level of items (See Table 13), and as a consequence, H_{02} was rejected.

Table 12. Shapiro-Wilk Normality Test Results for the Performances of the Students on Different Cognitive Level of Items

Groups	Statistic	*df*	*p*
Knowing	0.988	80	0.663
Applying	0.989	85	0.690
Reasoning	0.963	52	0.107

Table 13. One-Way Analysis of Variance for Performances of the Students on Different Cognitive Levels

Source	*SS*	*df*	*MS*	*F*
Between	7111.14	2	3555.57	14.12*
Within	53888.03	214	251.81	
Total	60999.17	216		

*$p < .05$

In order to evaluate pairwise differences among the means of performances on different cognitive level of items, Bonferroni test was performed and the

statistical significance was accepted at the $p < .05$ level. Bonferroni test has more power when the number of comparisons is small, whereas Tukey is more powerful when testing large numbers of means. Since it was aimed guaranteed control over the Type I error, Bonferroni test was used an appropriate post hoc test (Field, Miles, & Field, 2012).

This post hoc analysis revealed that the students' performance on knowing items ($M = 44.02$, $SD = 16.53$) is significantly higher than on applying ($M = 37.55$, $SD = 15.32$) and reasoning ($M = 29.01$, $SD = 15.72$) items. The students' performance on applying items ($M = 37.55$, $SD = 15.32$) is also significantly higher than on reasoning items ($M = 29.01$, $SD = 15.72$) (See Table 14).

Table 14. Bonferroni Test Results for Comparisons of the Students' Performance on Different Cognitive Levels

Group 1	Group 2	M_{Goup1} (%)	M_{Group2} (%)	p
Knowing	Applying	44.02	37.55	.029*
Knowing	Reasoning	44.02	29.01	.000*
Applying	Reasoning	37.55	29.01	.008*

*$p < .05$

H_{03}: There is no significant interaction effect between item format and cognitive domain of the items on students' performance in the TIMSS 2011 eighth grade mathematics assessment.

The descriptive statistics results of students' performance on MC and CR items across cognitive domains of the TIMSS eighth grade mathematics assessment are displayed in Table 15. The students' performance on MC items is higher on all cognitive level of items (knowing, applying, and reasoning) than their performance on CR items. In addition, from low to high cognitive level, the performances are decreasing in both MC and CR items. Hence the students' highest and lowest

performances are on MC knowing and CR reasoning items with 49.11% and 23.31%, respectively.

Table 15. Descriptive Statistics for Students' Performances on MC and CR Items across Cognitive Domains

Item Type	Number of Items	*M* (%)	*SD*	min.	max.
MC-Knowing	53	49.11	12.90	21.40	78.60
MC-Applying	47	43.92	12.30	19.70	70.90
MC-Reasoning	18	39.77	12.74	14.80	63.20
CR-Knowing	27	34.03	18.46	3.30	77.30
CR-Applying	38	29.68	15.14	1.70	60.10
CR-Reasoning	34	23.31	14.19	0.70	59.00

Two-way ANOVA test was conducted in order to compare the performances of the students on MC items and CR items across different cognitive levels. Prior to conducting the two-way ANOVA test, the normality of the data sets and the homogeneity of variances were assessed using Shapiro-Wilk and *Levene's* tests. Homogeneity of variances in the data sets ($p > .05$) was observed, and the data was normally distributed for each group of the sample except CR reasoning items ($p > .05$). When the box plot for students' performance on CR items across cognitive levels was examined, since the item M042066 fell 2.5 standard deviations from the mean of the students' performance on CR reasoning items, it was identified an outlier (See Figure 6). In order to verify normality assumption, the outlier was excluded from the analysis (Keppel & Wickens, 2004). Shapiro-Wilk normality test results after exclusion of the datum can be seen in Table 16. Homogeneity of variances in the new data set was also observed ($p = .068$).

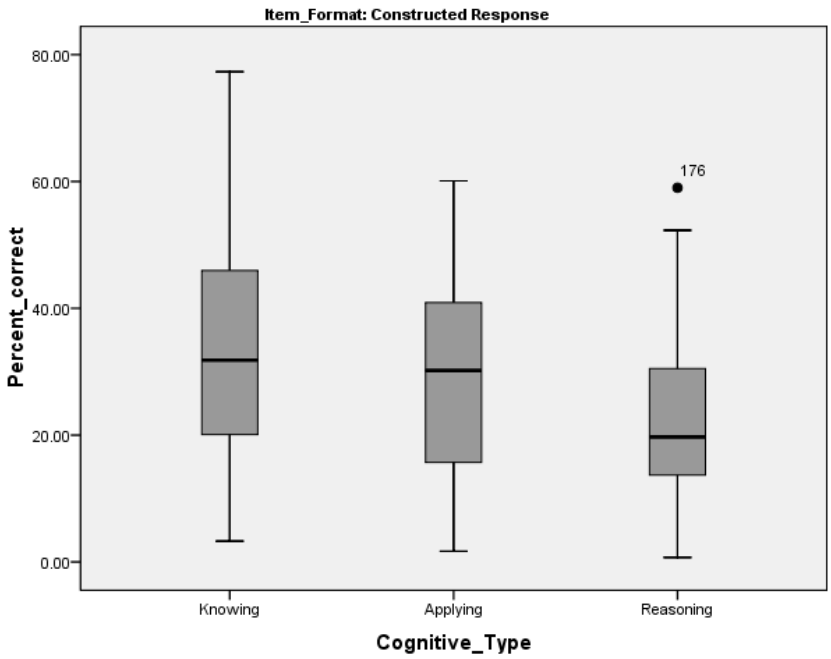


Figure 6 Box plot for students' performances on CR items across cognitive levels

Table 16. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Item Types

Groups	Statistic	*df*	*p*
MC Knowing	0.976	53	0.370
MC Applying	0.981	47	0.645
MC Reasoning	0.961	18	0.624
CR Knowing	0.974	27	0.703
CR Applying	0.977	38	0.605
CR Reasoning	0.936	33	0.051

The performance of the students was subjected to a two-way (2×3) ANOVA having two item format (MC, CR) and three cognitive levels of items (knowing, applying, and reasoning). All effects were statistically significant at the .05 significance level. This analysis revealed a significant main effect of item format, [$F(1,211) = 59.32$, $p = .00$, $\eta^2 = .220$], indicating that students' performance was significantly greater on MC items (M = 45.62, SD = 13.00) than on CR items (M = 28.68, SD = 16.22). The main effect of cognitive level of items was significant [$F(2,210) = 8.10$, $p = .00$, $\eta^2 = .072$]. However the interaction effect was non-

significant, [$F(2, 210) = 0.21$, $p = .81$, $\eta^2 = .002$], and as a consequence, H_{03} was not rejected (See Table 17). The main effects can be seen in Figure 7.

Table 17. Two-Way Analysis of Variance for Performances of the Students on MC and CR Items on Different Cognitive Levels

Source	*SS*	*df*	*MS*	*F*
Item Format	11598.24	1	11598.24	59.32*
Cognitive Level	3166.03	2	1583.01	8.10*
Item Format x Cognitive Level	82.64	2	41.32	0.21
Within Groups	41057.04	210	195.51	
Total	369037.25	216		

*$p < .05$

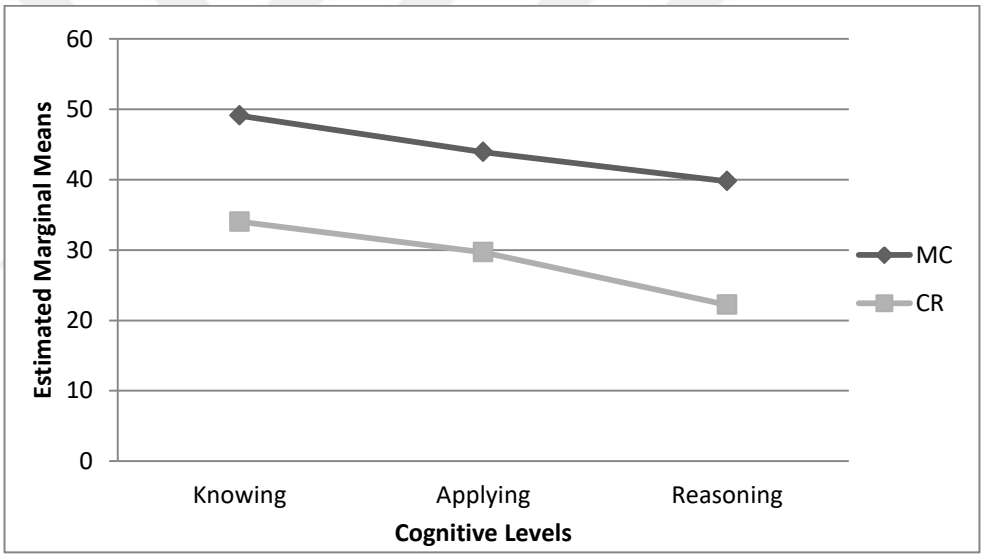


Figure 7 The effects of the interaction between item type and cognitive level of item on students' performance

4.3 Research question 3

What is the difference between Turkish students' performances on MC items and CR items in different content level of the TIMSS 2011 eighth grade mathematics assessment?

H_{01} was rejected in the part of research question 1.

H_{02}: There is no significant difference between students' performance in different content domain of items in the TIMSS 2011 eighth grade mathematics assessment.

Firstly, the performances of the Turkish students on number, algebra, geometry, and data and chance items were analyzed separately. It was seen that the percentage of the students' correctly answering data and chance items is the most with 46.13%. This is followed by performances on geometry, number and algebra items with 37.28%, 36.03%, and 34.81%, respectively. The descriptive statistics results of students' performances on four different content domains of the TIMSS eighth grade mathematics assessment can be seen in Table 18.

Table 18. Descriptive Statistics for the Performances of the Students on Number, Algebra, Geometry and Data and Chance Items

Item Type	Number of Items	*M* (%)	*SD*	min.	max.
Number	61	36.03	15.74	3.30	70.90
Algebra	70	34.81	17.19	17.90	77.30
Geometry	43	37.28	12.55	37.28	58.00
Data and Chance	43	46.13	19.09	0.70	78.60

A one-way ANOVA test was used in order to examine the performances of the students in four different content domains. The test for normality, examining standardized skewness and the Shapiro-Wilks test indicated that the data were normally distributed for each group of the sample (See Table 19). However, the *Levene's* test revealed that the homogeneity of variance assumption was not observed ($p = .04$). Therefore, the *Welch's* test was used (Huck, 2012). An alpha level of .05 was used for all subsequent analyses. The one-way ANOVA of students' performance on different content items revealed a statistically significant main effect, *Welch* $F(3, 110.19) = 3.63$, $p < .05$, indicating that not all content had the same

average of the students' performance, and as a consequence, this null hypothesis was rejected.

Table 19. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Content Domains

Groups	Statistic	*df*	*p*
Number	0.969	61	0.123
Algebra	0.975	70	0.183
Geometry	0.970	43	0.328
Data and Chance	0.974	43	0.428

Further, pairwise comparisons were performed using Games-Howell test (Huck, 2012). This post hoc test results are given in Table 20 and indicate that students' performance on data and chance items (M = 46.13, SD = 19.09) is significantly higher than their performance on number items (M = 36.03, SD = 15.74) as well as on algebra items (M = 34.81, SD = 17.19). The Cohen's d effect sizes for these two significant effects were 0.58 and 0.62, respectively. On the other hand the students' performances on other content domain items were similar.

Table 20. Games-Howell Post Hoc Results for Students' Performances on Different Content Items

Content domain	Mean Difference			
	Number	Algebra	Geometry	Data and Chance
Number	-	-	-	-
Algebra	-1.22	-	-	-
Geometry	1.25	2.47	-	-
Data and Chance	10.10*	11.32*	8.85	-

*$p < .05$

H_{03}: There is no significant interaction effect between item format and content domain of the items on students' performance in the TIMSS 2011 eighth grade mathematics assessment.

The descriptive statistics results of students' performance on MC and CR items across content domains of the TIMSS eighth grade mathematics assessment are shown in Table 21. The performance of the students on MC items is higher on all content domains (number, algebra, geometry, and data and chance) than the students' CR items performance. For MC items, the students' highest and lowest performances are on data and chance (M = 54.52) and geometry (M = 41.05) items, respectively. On the other hand, for CR items, the students' highest and lowest performances are on data and chance (M = 34.48) and algebra (M = 24.72) items, respectively.

Table 21. Descriptive Statistics for MC Items and CR Items in Different Content Domain

Item Type	Number of Items	*M* (%)	*SD*	min.	max.
MC-Number	31	44.27	10.90	14.80	70.90
MC-Algebra	37	43.82	11.79	19.70	65.70
MC-Geometry	25	41.05	11.20	15.70	58.00
MC-Data and Chance	25	54.52	15.16	28.60	78.60
CR-Number	30	27.53	15.56	3.30	60.40
CR-Algebra	33	24.72	16.80	3.50	77.30
CR-Geometry	18	32.06	12.75	10.30	57.40
CR-Data and Chance	18	34.48	18.14	0.70	60.10

In order to compare the performances of the students on MC items and CR items at different content domains, two-way ANOVA was conducted. Prior to conducting the two-way ANOVA test, the normality of the data sets and the homogeneity of variances were assessed using Shapiro-Wilk and *Levene's* tests. Homogeneity of variances in the data sets (p = .055) was not observed, and the data was normally distributed for each group of the sample except CR algebra items (p > .05). When the box plots for students' performance on MC and CR items across cognitive levels were examined, it was seen extreme scores on MC number and CR algebra items (See Figure 8 and Figure 9). Since the items M032662 and M042041 fell 2.7 and 2.4 standard deviations away from the mean of the students' performance

on MC number items, they were obviously outliers. Although MC number items are normally distributed and the outliers are seem non problematic, they enlarge the variance. Also, due to the items M042198A and M042066 fell 3.1 and 2.0 standard deviations away from the mean of the students' performance on CR algebra items, they were outliers. In order to verify normality assumption the outliers were excluded from the analysis (Keppel & Wickens, 2004). However the violation of normality assumption for CR algebra items was not resolved. Then another extreme item M042228 which falls 1.6 standard deviation from the mean of the students' performance on CR algebra items was excluded and the assumption was satisfied. After exclusion of the datum, Shapiro-Wilk normality test results can be seen in Table 22. The *Levene's* test revealed that the homogeneity of variance assumption was not observed (p = .002). After the variances for each sample group were examined, it was seen that only MC data and chance and CR data and chance items validate the assumption. In order to verify homogeneity of variances assumption, the nominal significance level is reduced and an alpha level of .001 was used for all subsequent analyses (Keppel & Wickens, 2004).

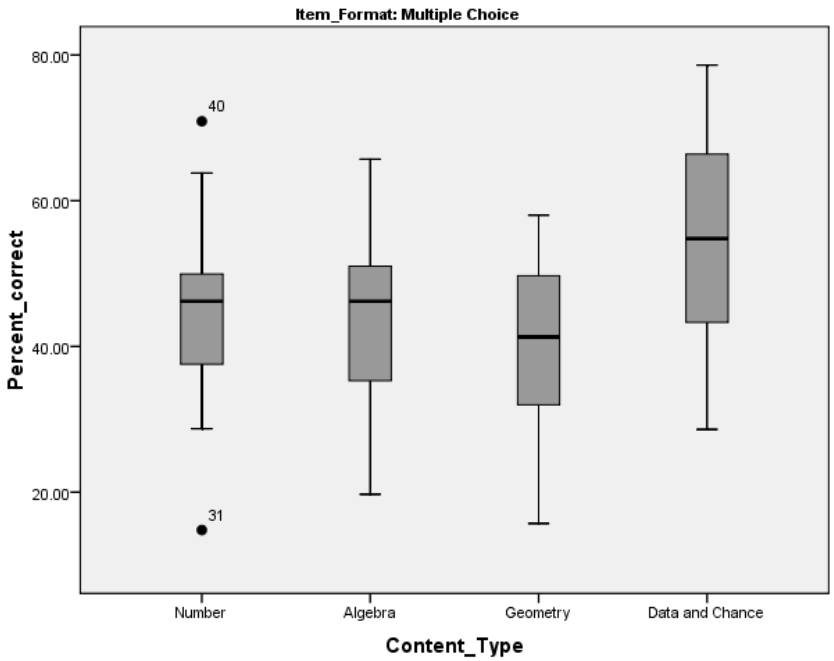


Figure 8 Box plot for students' performances on MC items across content domains

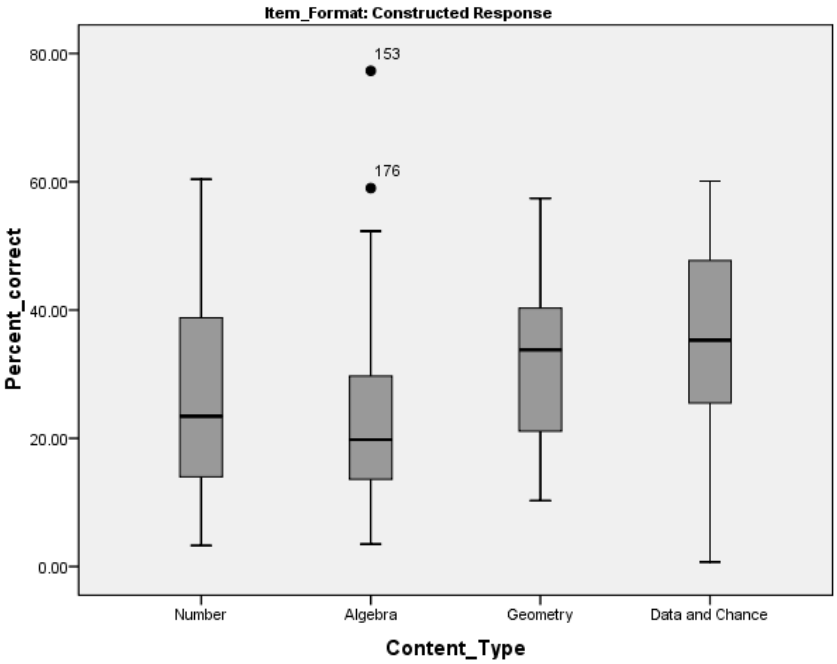


Figure 9 Box plot for students' performances on CR items across content domains

Table 22. Shapiro-Wilk Normality Test Results for the-Performances of the Students on Different Item Types

Groups	Statistic	*df*	*p*
MC Number	0.977	29	0.747
MC Algebra	0.976	37	0.605
MC Geometry	0.950	25	0.247
MC Data and Chance	0.953	25	0.286
CR Number	0.945	30	0.122
CR Algebra	0.931	30	0.053
CR Geometry	0.971	18	0.824
CR Data and Chance	0.951	18	0.440

The performance of the students was subjected to a two-way (2×4) ANOVA having two item format (MC, CR) and four content domains of items (number, algebra, geometry, and data and chance). All effects were statistically significant at the .025 significance level. This analysis revealed a significant main effect of item format, [$F(1, 204) = 86.60$, $p = .00$, $\eta^2 = .274$], indicating that students' performance was significantly greater on MC items (M = 45.62, SD = 13.00) than on CR items (M = 28.68, SD = 16.22). The main effect of cognitive level of items was significant [$F(3, 204) = 7.44$, $p = .00$, $\eta^2 = .130$]. However the interaction effect was non-significant, [$F(3, 204) = 2.54$, $p > .025$, $\eta^2 = .036$], and as a consequence, H_{03} was not rejected (See Table 23). The main effects can be seen in Figure 10.

Table 23. Two-Way Analysis of Variance for Performances of the Students on MC and CR Items on Different Content Domains

Source	*SS*	*df*	*MS*	*F*
Item Format	14801.64	1	14801.64	86.60*
Content Domain	3815.14	3	1271.71	7.44*
Item Format x Content Domain	1304.26	3	434.75	2.54
Within Groups	34869.23	204	170.93	
Total	355215.11	212		

*$p < .001$

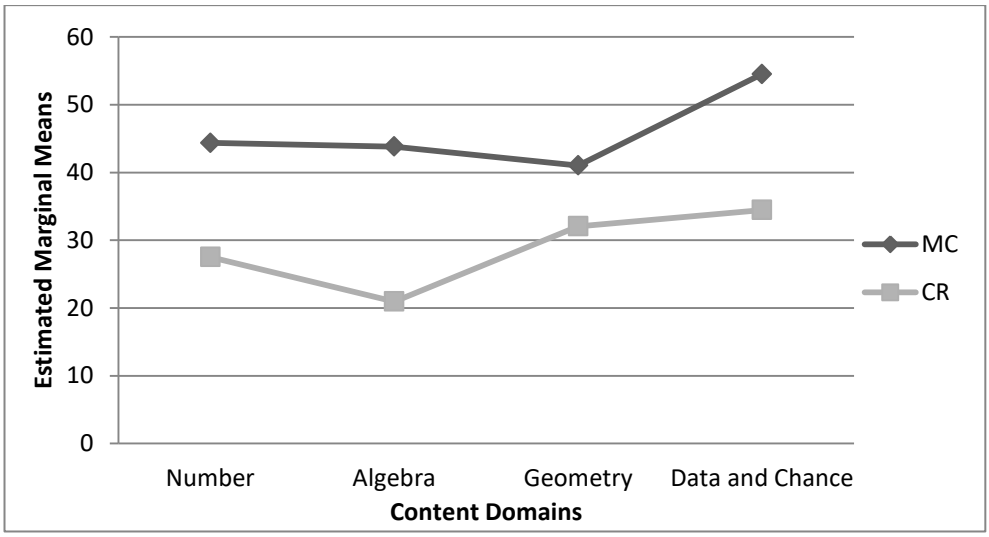


Figure 10 The effects of the interaction between item type and content domain of item on students' performance

4.4 Research question 4

What is the difference between countries rank when looking at students' performances on different item type (item format, cognitive and content domains) of the TIMSS 2011 eighth grade mathematics assessment?

H_0: There is no significant difference in ranking of Turkey compared to other countries who attended the TIMSS 2011 eighth grade mathematics assessment when different item formats are used.

Forty two countries participated in the TIMSS 2011 eighth grade mathematics assessment. The TIMSS 2011 eighth grade mathematics achievement booklets include totally 217 mathematics items; however, some countries did not answer all questions. The list of the items that were not included in country specific booklets is presented in Table 24. The analyses for these countries were conducted excluding the items for each specific country.

Table 24. The Items were not included to the Analysis of the Country

Country	Item	Item format	Cognitive domain	Content domain
Chile	M0323331	MC	Applying	Geometry
Macedonia	M042158	MC	Knowing	Data and Chance
Norway	M052173	MC	Applying	Algebra
Romania	M042261	MC	Knowing	Data and Chance
Sweden	M042103	CR	Knowing	Algebra
Sweden	M032538	CR	Knowing	Algebra

For each country, average percent correct of students were calculated for MC, CR; knowing, applying, reasoning; number, algebra, geometry, data and chance; MC knowing, MC applying, MC reasoning, CR knowing, CR applying, CR reasoning; MC number, MC algebra, MC geometry, MC data and chance, CR number, CR algebra, CR geometry, CR data and chance items of the TIMSS 2011 eighth grade mathematics assessment separately. Average percent correct of the students from 42 countries across different item type in the TIMSS 2011 is displayed in Table A5 (See Appendix D). For each item type, all countries were ranked based on their average percent correct, separately. The rankings based on average percent correct for each item type are showed in Table A6 (See Appendix D). Countries are sorted in alphabetical order in both Table A5 and Table A6. According to the average percent correct ranking based on all (overall) items, Korea (M = 72.85) is at the highest rank and Ghana (M = 19.06) is the last. The international average is 40.65% and Turkey (M = 37.89) is at the 23rd rank and below the international average.

When we looked the overall average percent correct ranking with the rankings based on different item types, moving up and down the rankings were observed. The overall average percent correct ranking and the average percent correct rankings based on different item types were compared by using Kendall's tau-b test which is a nonparametric procedure to determine the strength and direction of the association between two ordinal variables (Rovai, Baker, & Ponton, 2013).

There is a strong, positive correlation between the overall average percent correct ranking and the average percent correct ranking based on MC items which is statistically significant (τ_b = .965, p = .000). Also, there is a strong and positive correlation between the overall average percent correct ranking and the average percent correct ranking based on CR items which is statistically significant (τ_b = .972 p = .000). In addition, the country rankings based on average percent correct of different item types are significantly similar to the overall average percent correct ranking, and H_0 was rejected. All Kendall's tau-b test results can be seen in Table 25.

Table 25. Kendall's tau-b Test Results

Item Type	Kendall's tau-b Coefficient	p
MC	0.965	0.000
CR	0.972	0.000
Knowing	0.944	0.000
Applying	0.949	0.000
Reasoning	0.926	0.000
Number	0.940	0.000
Algebra	0.812	0.000
Geometry	0.837	0.000
Data and chance	0.798	0.000
MC Knowing	0.921	0.000
MC Applying	0.937	0.000
MC Reasoning	0.875	0.000
CR Knowing	0.916	0.000
CR Applying	0.912	0.000
CR Reasoning	0.902	0.000
MC Number	0.912	0.000
MC Algebra	0.789	0.000
MC Geometry	0.849	0.000
MC Data and chance	0.849	0.000
CR Number	0.923	0.000
CR Algebra	0.856	0.000
CR Geometry	0.816	0.000
CR Data and chance	0.791	0.000

The position of Turkey based on average percent correct ranking on different item type changes between 18 to 26th rank. As it can be seen in Table 26, when overall items ranking is compared with MC and CR items rankings, there are 25 countries which remain stable, there are 11 countries change their position by one, 3 countries by two, and another 3 countries by three on MC items ranking, while there

are 31 countries remain stable, 5 countries change their position by one, another 5 countries by two, and one country by three on CR items ranking.

Table 26. Position Changes on Overall Items Compared with MC and CR Items

Country	Overall	MC	CR	Difference between overall and MC items ranks	Difference between overall and CR items ranks
Korea	1	1	1	0	0
Singapore	2	2	2	0	0
Chinese Tai.	3	3	3	0	0
Hong Kong	4	4	4	0	0
Japan	5	5	5	0	0
Russian Fed.	6	6	6	0	0
Israel	7	7	7	0	0
Finland	8	8	8	0	0
United States	9	9	11	0	-2
England	10	10	10	0	0
Hungary	11	11	9	0	2
Australia	12	12	12	0	0
Slovenia	13	14	13	-1	0
Lithuania	14	15	14	-1	0
Italy	15	13	17	2	-2
New Zealand	16	16	16	0	0
Kazakhstan	17	18	15	-1	2
Ukraine	18	17	19	1	-1
Sweden	19	19	18	0	1
Norway	20	23	20	-3	0
Armenia	21	20	21	1	0
Romania	22	22	22	0	0
Turkey	23	21	23	2	0
U. Arab E.	24	24	24	0	0
Lebanon	25	26	25	-1	0
Malaysia	26	25	26	1	0
Georgia	27	27	28	0	-1
Macedonia	28	28	27	0	1
Thailand	29	29	29	0	0
Qatar	30	30	31	0	-1
Iran	31	31	33	0	-2
Bahrain	32	34	32	-2	0
Tunisia	33	36	30	-3	3
Palestinian	34	33	34	1	0
Jordan	35	32	35	3	0
Chile	36	35	36	1	0
Saudi Arabia	37	37	37	0	0
Syrian A. R.	38	38	38	0	0
Oman	39	40	39	-1	0
Indonesia	40	39	40	1	0
Morocco	41	41	41	0	0
Ghana	42	42	42	0	0

The ranking position changes of Turkey was analyzed in detail by using Dunnett's test which is used to compare each treated group with the same control group (Huck, 2012). By the help of this post hoc test, the mean of the performances of Turkey and the other countries were compared whether there was a noticeable difference or not. The results can be seen in Table A7 (See Appendix D). Table 27 shows the means of the performance of the countries while Table 28 displays the rankings based on these means. Moreover, the countries are categorized in terms of the significance of the difference between each average and the average of Turkey. In Table 27 and Table 28, the bold ones are not significantly different from the performance of Turkey, also the countries are sorted in overall average percent correct ranking. When the Table 27 is examined, it stands out that the mean of the performance of Turkey in each item type is not significantly different from the international average according to Dunnett's test results (See Appendix F). More specifically, the overall performance of Turkey is below the international average but there was no significant difference between them. The overall performances of Kazakhstan, Ukraine, Sweden, Norway, Armenia, and Romania were above while United Arab Emirates was below the overall performance of Turkey. However, there were no significant differences between the performances of all seven countries and Turkey. In addition, the performance of Turkey are only above the international average on data and chance, MC data and chance items, and CR data and chance items, but there was no significant difference between the performance of Turkey and international average. Moreover, the highest performance of Turkey was on MC data and chance items (M = 54.52) and the lowest performance on CR reasoning items (M = 23.31) which were not significantly different from international averages.

Table 27. Average Percentages Correct at Eighth Grade for 42 Countries across Different Item Type in TIMSS

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Korea	72.85	78.14	66.55	79.84	73.31	61.36	77.29	69.99	71.50	72.56	83.44	77.67	65.23	72.99	667.91	59.31	80.70	75.71	70.70	82.83	71.22	60.49	72.62	58.31
Singapore	72.68	77.90	66.46	81.48	72.46	59.52	77.39	70.95	71.61	69.91	82.53	75.98	66.49	77.50	68.11	55.83	81.40	75.55	72.03	77.55	68.70	61.73	71.02	59.29
Chinese Tai.	70.89	76.71	63.96	77.17	71.94	59.54	72.05	71.37	73.04	66.34	80.36	74.78	70.18	70.11	68.42	53.91	77.01	76.18	75.00	75.29	64.51	63.19	70.32	53.91
Hong Kong	66.98	73.32	59.42	75.56	67.27	53.32	70.68	63.02	69.36	65.80	80.02	69.65	63.14	67.20	64.32	48.12	75.39	69.74	69.43	73.92	62.26	50.98	69.26	54.52
Japan	63.23	68.84	56.55	69.26	64.09	52.57	62.94	59.50	66.32	66.64	75.00	67.40	58.99	63.08	60.00	49.17	66.68	67.01	65.11	74.86	56.83	49.11	68.01	55.22
Russian Fed.	55.65	62.93	46.99	66.83	54.42	40.47	58.63	56.03	54.28	**52.19**	70.53	57.86	52.17	58.71	50.17	34.27	64.57	65.76	54.92	**59.40**	48.57	40.03	53.39	**42.18**
Israel	50.60	57.56	42.29	59.85	49.48	38.20	54.68	48.29	**45.20**	**53.95**	63.89	54.54	**46.54**	53.22	43.21	33.78	59.04	56.53	**47.74**	**62.05**	46.23	36.01	**41.68**	**42.71**
Finland	49.01	56.32	40.31	57.60	49.28	**35.37**	55.57	**39.06**	45.83	59.10	64.16	53.18	**46.69**	50.80	44.46	**29.37**	57.52	**48.77**	**48.78**	68.46	49.68	**25.56**	**41.74**	**46.09**
United States	48.32	56.24	38.89	60.79	45.84	**33.21**	53.17	43.11	**41.85**	**56.41**	64.70	51.33	**42.93**	51.72	**39.05**	**28.06**	56.96	54.96	**45.41**	**64.42**	44.33	**29.82**	**36.91**	**45.29**
England	48.23	55.30	39.81	56.80	47.79	**35.77**	52.44	**38.99**	46.18	59.36	61.63	51.54	**45.52**	**45.14**	43.16	**30.61**	56.24	**48.06**	**49.56**	67.96	44.89	**28.82**	**41.49**	**47.43**
Hungary	48.10	54.78	40.14	58.46	46.49	**34.81**	52.56	**41.14**	47.29	**53.91**	64.53	**49.42**	**44.06**	50.94	42.87	**29.91**	55.39	**47.61**	**50.48**	**61.40**	45.41	**28.87**	**42.86**	**43.50**
Australia	47.46	54.75	38.77	56.68	46.39	**35.01**	51.93	**38.33**	45.79	57.64	63.45	**50.72**	**46.03**	50.21	41.03	**29.18**	55.11	**45.61**	**49.66**	67.01	44.09	**27.05**	**40.41**	**44.63**
Slovenia	46.64	54.67	37.07	57.01	45.08	**33.25**	51.83	**38.17**	46.36	**53.37**	64.57	**48.94**	**45.07**	**46.64**	40.30	**26.99**	56.18	**46.90**	**48.42**	**63.41**	42.53	**24.04**	**43.49**	**39.42**
Lithuania	46.28	54.03	37.05	56.22	46.28	**31.00**	48.68	**40.16**	46.25	**52.89**	62.66	**50.72**	**42.16**	**49.38**	40.79	**25.10**	**51.76**	**49.98**	**49.67**	**61.46**	41.03	**25.60**	**41.49**	**40.98**
Italy	45.75	54.68	**35.10**	54.31	45.57	**32.86**	48.84	**38.64**	48.54	**50.14**	60.98	51.70	**47.02**	**45.58**	**38.00**	**25.37**	56.83	**48.28**	52.17	**59.40**	**36.59**	**24.92**	**43.49**	**37.29**
New Zealand	43.67	**50.40**	35.65	51.53	**43.26**	**32.23**	47.33	**34.87**	**42.09**	**54.38**	59.17	**47.13**	**41.18**	**44.88**	**38.48**	**27.50**	**50.22**	**41.96**	**45.50**	**63.39**	41.05	**24.86**	**37.37**	**41.86**
Kazakhstan	**42.82**	**48.40**	**36.16**	**52.58**	**41.18**	**30.46**	**44.38**	**43.32**	**43.31**	**39.29**	**54.36**	**44.45**	**39.13**	**46.83**	**37.15**	**25.87**	**49.09**	**51.21**	**42.18**	**44.92**	**34.64**	**29.74**	44.88	**31.47**
Ukraine	**41.62**	**49.86**	**31.80**	51.44	**40.51**	**28.34**	**42.68**	**39.53**	**41.05**	**44.09**	56.99	**45.89**	**40.17**	**41.18**	**33.86**	**22.07**	**49.78**	**51.08**	**43.57**	**52.52**	**32.09**	**26.58**	**37.54**	**32.38**
Sweden	**41.25**	**47.61**	**33.52**	**49.84**	**41.33**	**28.24**	49.09	**31.24**	**35.61**	**51.60**	**54.53**	**45.19**	**37.36**	**45.46**	**36.56**	**23.41**	**50.87**	**37.86**	**40.82**	**59.21**	43.26	**21.05**	**28.37**	**41.04**
Norway	**39.00**	**45.16**	**31.73**	**46.22**	**39.37**	**27.31**	**45.99**	25.69	**36.14**	**53.32**	**52.26**	**43.92**	**35.71**	**43.58**	**33.87**	**22.86**	**46.57**	33.09	**40.72**	**61.10**	40.58	**16.95**	**29.77**	**42.51**
Armenia	**38.90**	**45.80**	**30.68**	**50.18**	**36.46**	**25.55**	**44.11**	**41.84**	**35.41**	30.23	**50.73**	**41.15**	**35.97**	**42.63**	**20.89**	**20.03**	**50.75**	**50.70**	**35.28**	37.56	**32.66**	**26.41**	**35.59**	**20.05**
Romania	**37.99**	**45.45**	**29.17**	**47.27**	**36.15**	**26.89**	**38.63**	**38.40**	**36.87**	**37.51**	**51.59**	**41.46**	**37.42**	**38.05**	**29.59**	**21.31**	**45.29**	**46.56**	**39.40**	**45.15**	**28.11**	**24.61**	**33.35**	**27.34**
Turkey	**37.89**	**45.62**	**28.68**	**44.02**	**37.55**	**29.01**	**36.03**	**34.81**	**37.28**	**46.13**	**49.11**	**43.92**	**39.77**	**34.03**	**29.68**	**23.31**	**44.27**	**43.82**	**41.05**	**54.52**	**27.53**	**24.72**	**32.06**	**34.48**
U. Arab E.	**36.78**	**44.89**	**27.12**	**48.30**	**33.64**	**24.20**	**40.94**	**34.06**	**32.55**	**39.55**	**52.28**	**39.83**	**33.78**	**37.89**	**25.97**	**19.14**	**48.24**	**44.22**	**36.54**	**46.78**	**28.03**	**22.68**	**27.01**	**29.50**
Lebanon	34.74	**43.19**	**24.68**	**46.75**	**31.76**	21.14	**38.02**	**35.51**	**33.01**	30.57	**50.29**	**36.82**	**32.94**	**34.68**	**25.49**	**14.89**	**46.46**	**45.18**	**34.03**	37.53	**23.18**	**18.81**	**31.59**	**20.90**
Malaysia	34.23	**43.20**	**23.54**	**43.30**	**32.91**	**22.44**	**40.04**	27.74	**34.37**	**36.42**	**48.69**	**39.53**	**36.13**	**32.93**	**24.72**	**15.19**	**48.26**	**37.08**	**39.17**	44.61	**25.92**	**14.49**	**27.71**	**25.03**
Georgia	33.47	**42.47**	22.74	**43.16**	**31.77**	**21.33**	**35.69**	**33.59**	**30.37**	33.21	**48.98**	**38.57**	**32.89**	**30.95**	**23.37**	**15.21**	**41.66**	**44.21**	**35.23**	**43.07**	**25.28**	**18.35**	**23.62**	**19.51**
Macedonia	32.79	40.74	**23.40**	**41.04**	31.03	**23.15**	**32.86**	**32.88**	**32.85**	32.49	**46.20**	**37.74**	**33.44**	**31.59**	**22.72**	**17.70**	**40.42**	**43.22**	**36.04**	40.54	**22.01**	**21.28**	**28.41**	**21.76**
Thailand	31.14	39.38	21.32	**38.34**	30.24	**21.53**	**33.37**	**26.82**	**29.57**	**36.57**	**43.78**	36.73	**32.29**	**25.38**	**22.21**	**15.83**	**41.74**	**36.37**	**33.44**	**45.20**	**21.27**	**16.02**	**24.20**	**24.58**
Qatar	30.38	38.78	20.36	**39.20**	28.23	20.31	**32.75**	**28.65**	27.40	32.79	**45.10**	35.11	**30.39**	**28.84**	19.73	**14.97**	**40.91**	**37.28**	**31.73**	41.08	**20.46**	**16.40**	**21.39**	**21.27**
Iran	30.01	38.55	19.82	**36.52**	28.74	**22.04**	**30.21**	26.87	**32.90**	31.92	**43.94**	35.66	**33.62**	**25.45**	**20.20**	**15.91**	**39.29**	**36.70**	**35.73**	40.91	**17.97**	**14.69**	**28.97**	**19.44**
Bahrain	29.51	37.29	20.24	**37.34**	27.60	20.59	**29.98**	26.86	27.22	35.46	**43.19**	34.44	**30.09**	**29.72**	19.14	**15.56**	**38.24**	**35.42**	**31.38**	41.72	**18.57**	**14.90**	**21.44**	**26.77**
Tunisia	29.31	36.42	20.84	37.20	28.24	18.93	**32.65**	24.91	**29.70**	31.35	40.54	33.66	**28.21**	**26.31**	**21.54**	14.02	**38.22**	**35.30**	**32.35**	36.72	**21.42**	13.27	**26.02**	**23.89**
Palestinian	29.10	37.47	19.13	**36.98**	27.36	19.85	**30.68**	26.63	**30.39**	29.60	**41.61**	34.04	**32.03**	**27.68**	19.09	13.40	**39.50**	**35.99**	**33.03**	36.88	**16.66**	13.96	**26.72**	**19.49**
Jordan	29.04	37.53	18.92	**37.05**	26.91	20.21	28.40	**29.02**	28.41	30.62	**43.48**	33.69	**29.86**	**25.37**	18.52	**15.10**	**36.98**	**38.53**	31.33	38.37	16.22	**15.64**	**24.35**	**19.85**
Chile	28.55	36.89	18.69	**33.81**	28.91	19.87	**29.89**	22.30	29.02	**36.34**	**41.91**	35.53	**32.73**	**25.28**	**20.89**	13.05	**37.56**	30.68	**34.83**	**46.08**	**20.79**	12.34	**21.27**	**22.82**
Saudi Arabia	26.76	35.83	15.94	**35.04**	24.68	17.40	28.86	23.79	24.36	31.01	40.78	33.17	**29.43**	**26.90**	14.19	11.03	**38.76**	32.72	30.75	38.65	15.13	11.94	15.48	**20.39**
Syrian A. R.	25.06	33.41	15.09	31.19	24.32	16.82	25.22	24.17	25.97	25.34	36.35	31.88	28.09	**21.56**	14.97	10.85	**34.17**	**34.65**	29.88	32.19	13.49	10.82	**20.54**	15.82
Oman	24.53	32.99	14.45	31.40	22.91	16.64	24.15	23.11	25.22	26.72	37.42	30.03	27.77	**21.04**	14.09	10.75	**33.97**	31.81	29.40	33.60	11.11	11.17	19.41	17.17
Indonesia	24.45	33.15	14.07	30.76	23.33	16.55	24.68	21.80	24.70	28.17	38.00	29.92	**29.10**	**18.89**	15.18	9.90	**34.04**	31.11	30.25	34.95	12.45	9.73	16.99	**18.76**
Morocco	22.69	30.65	13.21	28.25	22.65	14.23	24.14	19.63	25.05	23.27	33.85	28.67	26.48	17.73	15.19	7.74	31.35	29.25	29.28	30.46	13.01	8.04	19.18	13.28
Ghana	19.06	27.89	8.52	25.08	17.68	12.05	19.28	18.58	17.77	20.80	30.41	25.76	24.53	14.53	7.69	5.44	29.11	27.78	23.79	27.98	6.61	6.66	9.40	10.83
Inter. avg.	**40.65**	**48.22**	**31.64**	**49.08**	**39.62**	**29.37**	**43.30**	**36.98**	**39.71**	**43.81**	**53.95**	**44.97**	**39.83**	**39.53**	**33.01**	**23.83**	**51.25**	**46.94**	**42.75**	**51.82**	**35.10**	**25.82**	**35.50**	**32.69**

Bold ones are not significantly different with Turkey (Dunnett's test $p < 0.05$).

Table 28. Rankings Based on Average Percent Correct for each Item Type

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Korea	1	1	1	2	1	1	2	3	3	1	1	1	3	2	1	1	2	2	3	1	1	3	1	2
Singapore	2	2	2	1	2	3	1	2	2	2	2	2	2	1	3	2	1	3	2	2	2	2	2	1
Chinese Tai.	3	3	3	3	3	2	3	1	1	4	3	3	1	3	2	3	3	1	1	3	3	1	3	5
Hong Kong	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	4	4	4	5	4	4	4	4
Japan	5	5	5	5	5	5	5	5	5	3	5	5	5	5	5	4	5	5	5	4	5	5	5	3
Russian Fed.	6	6	6	6	6	6	6	6	6	**16**	6	6	6	6	6	6	6	6	6	**16**	7	6	6	**13**
Israel	7	7	7	8	7	7	8	7	**14**	**11**	11	7	**9**	7	8	7	7	7	**14**	**12**	8	7	**12**	**11**
Finland	8	8	8	10	8	**9**	7	**14**	12	7	10	8	**8**	10	7	**10**	8	**13**	**12**	6	6	**16**	**11**	**7**
United States	9	9	11	7	13	**13**	9	9	**17**	**9**	7	11	**14**	8	**14**	**12**	9	8	**16**	**9**	11	**8**	**18**	**8**
England	10	10	10	12	9	**8**	11	**15**	11	6	14	10	**11**	**17**	9	**8**	11	**15**	**11**	7	10	**11**	**14**	**6**
Hungary	11	11	9	9	10	**11**	10	**11**	8	**12**	9	**14**	**13**	9	10	**9**	13	**16**	**8**	**14**	9	**10**	**10**	**10**
Australia	12	12	12	13	11	**10**	12	**18**	13	8	12	**12**	**10**	11	11	**11**	14	**19**	**10**	8	12	**12**	**15**	**9**
Slovenia	13	14	13	11	15	**12**	13	**19**	9	**13**	8	**15**	**12**	**14**	13	**14**	12	**17**	**13**	**10**	14	**21**	**9**	**17**
Lithuania	14	15	14	14	12	**16**	16	**12**	10	**15**	13	**13**	**15**	**12**	12	**17**	**15**	**12**	**9**	**13**	16	**15**	**13**	**16**
Italy	15	13	**17**	15	14	**14**	15	**16**	7	**18**	15	9	**7**	**15**	**16**	**16**	10	**14**	7	**17**	**18**	**17**	**8**	**18**
New Zealand	16	**16**	16	17	**16**	**15**	17	**21**	**16**	**10**	16	**16**	**16**	**18**	**15**	**13**	**18**	**25**	**15**	**11**	15	**18**	**17**	**14**
Kazakhstan	**17**	**18**	**15**	**16**	**18**	**17**	**19**	**8**	**15**	**22**	**19**	**19**	**19**	**13**	**17**	**15**	**20**	**9**	**18**	**25**	**19**	**9**	7	**21**
Ukraine	**18**	**17**	**19**	18	**19**	**19**	**21**	**13**	**18**	**20**	17	**17**	**17**	**21**	**20**	**21**	**19**	**10**	**17**	**20**	**21**	**13**	**16**	**20**
Sweden	**19**	**19**	**18**	**20**	**17**	**20**	14	**26**	**22**	**17**	**18**	**18**	**21**	**16**	**18**	**18**	**16**	**27**	**20**	**18**	13	**24**	**26**	**15**
Norway	**20**	**23**	**20**	**24**	**20**	**21**	**18**	34	**21**	**14**	**21**	**21**	**24**	**19**	**19**	**20**	**23**	36	**21**	**15**	17	**27**	**23**	**12**
Armenia	**21**	**20**	**21**	**19**	**22**	**23**	**20**	**10**	**23**	36	**23**	**23**	**23**	**20**	**30**	**23**	**17**	**11**	**27**	34	**20**	**14**	**19**	**33**
Romania	**22**	**22**	**22**	**22**	**23**	**22**	**24**	**17**	**20**	**23**	**22**	**22**	**20**	**22**	**22**	**22**	**25**	**18**	**22**	**24**	**22**	**20**	**20**	**23**
Turkey	**23**	**21**	**23**	**25**	**21**	**18**	**26**	**22**	**19**	**19**	**25**	**20**	**18**	**25**	**21**	**19**	**26**	**23**	**19**	**19**	**24**	**19**	**21**	**19**
U. Arab E.	**24**	**24**	**24**	**21**	**24**	**24**	**22**	**23**	**28**	**21**	**20**	**24**	**25**	**23**	**23**	**24**	**22**	**21**	**24**	**21**	**23**	**22**	**28**	**22**
Lebanon	25	**26**	**25**	**23**	**27**	30	**25**	**20**	**25**	35	**24**	**28**	**28**	**24**	**24**	**33**	**24**	**20**	**30**	35	**27**	**25**	**22**	**31**
Malaysia	26	**25**	**26**	**26**	**25**	**26**	**23**	29	**24**	**25**	**27**	**25**	**22**	**26**	**25**	**30**	**21**	**29**	**23**	26	**25**	**33**	**27**	**25**
Georgia	27	**27**	28	**27**	**26**	**29**	**27**	**24**	**30**	28	**26**	**26**	**29**	**28**	**26**	**29**	**28**	**22**	**28**	**27**	**26**	**26**	**33**	**35**
Macedonia	28	28	**27**	**28**	28	**25**	**29**	**25**	**27**	30	**28**	**27**	**27**	**27**	**27**	**25**	**30**	**24**	**25**	31	**28**	**23**	**25**	**29**
Thailand	29	29	29	**30**	29	**28**	**28**	**32**	**32**	**24**	**31**	29	**31**	**35**	**28**	**27**	**27**	**31**	**31**	**23**	**30**	**29**	**32**	**26**
Qatar	30	30	31	**29**	33	32	**30**	**28**	35	29	**29**	32	**33**	**30**	33	**32**	**29**	**28**	**34**	29	**32**	**28**	**35**	**30**
Iran	31	31	33	**35**	31	**27**	**33**	**30**	**26**	31	**30**	30	**26**	**34**	**32**	**26**	**32**	**30**	**26**	30	**34**	**32**	**24**	**37**
Bahrain	32	34	32	**31**	34	31	**34**	31	36	27	**33**	33	**34**	**29**	34	**28**	**34**	**33**	**35**	28	**33**	**31**	**34**	**24**
Tunisia	33	36	30	32	32	36	**31**	35	**31**	32	37	36	**38**	**33**	**29**	34	**35**	**34**	**33**	37	**29**	35	**30**	**27**
Palestinian	34	33	34	**34**	35	35	**32**	33	**29**	37	**35**	34	**32**	**31**	35	35	**31**	**32**	**32**	36	**35**	34	**29**	**36**
Jordan	35	32	35	**33**	36	33	37	**27**	34	34	**32**	35	**35**	**36**	36	**31**	**37**	**26**	36	33	36	**30**	**31**	**34**
Chile	36	35	36	**37**	30	34	**35**	39	33	**26**	**34**	31	**30**	**37**	**31**	36	**36**	40	**29**	**22**	**31**	36	**36**	**28**
Saudi Arabia	37	37	37	**36**	37	37	36	37	41	33	36	37	**36**	**32**	40	37	**33**	37	37	32	37	37	41	**32**
Syrian A. R.	38	38	38	39	38	38	38	36	37	40	40	38	39	**38**	39	38	**38**	**35**	39	40	38	39	**37**	40
Oman	39	40	39	38	40	39	40	38	38	39	39	39	40	**39**	41	39	**40**	38	40	39	41	38	38	39
Indonesia	40	39	40	40	39	40	39	40	40	38	38	40	**37**	**40**	38	40	**39**	39	38	38	40	40	40	**38**
Morocco	41	41	41	41	41	41	41	41	39	41	41	41	41	41	37	41	41	41	41	41	39	41	39	41
Ghana	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42

Bold ones are not significantly different with Turkey (Dunnett's test $p < 0.05$).

Besides, when we looked the reasoning and MC reasoning items on which orders Turkey is at the highest ranks, it is deduced that the performance of Turkey is below the international average, still there was no significant difference between them (See Table 28). Similarly, Turkey is at the lowest rank on number and MC number items. Although the performance of Turkey on number and on MC number items are below the international averages, there was no significant difference between them (See Table 28).

Further, we looked the ranking position changes of Turkey on different item type in detail. It is noticeable in Table 29 that, the rank of Turkey is higher on MC items than on overall items and CR items. When the ranking position of Turkey on overall items was compared with MC items, it was seen that the position of Turkey increases by two ranks. Turkey (M = 45.62) passes Romania (M = 45.16) and Norway (M = 45.45) on MC items, however there are no significant difference between the performance of the students from Norway and Turkey (p = 1.000) and from Romania and Turkey (p = 1.000), according to Dunnett's test results. On the other hand, the ranking position of Turkey on CR items remained stable. The rank of Turkey is highest (18th rank) on reasoning and MC reasoning items, passes five countries in each ranking, but there are no significant differences between the performances of the students. On the other hand, the lowest rank of Turkey (26th rank) is on the ranking of number and MC number items. Turkey falls down United Arab Emirates, Lebanon and Malaysia, yet there are no significant differences between the performances of the students.

Finally, Dunnett's test was conducted to compare international average and the mean of the performances of other countries. In Table 30, noticeable differences between international average and the performances of the countries can be seen. It is

Table 29. The Ranking Position Changes of Turkey on Different Item Type Compared with Overall Ranking[1]

Item Type	Rank of Turkey	Difference with overall rank	Turkey passes or falls down…	Significantly difference[2]
MC	21	+2	Romania, Norway	None[3]
CR	23	0		None
Knowing	25	-2	United Arab Emirates, Lebanon	None
Applying	21	+2	Armenia, Romania	None
Reasoning	18	+5	Ukraine, Sweden, Norway, Armenia, Romania	None
Number	26	-3	United Arab Emirates, Lebanon and Malaysia	None
Algebra	22	+1	Sweden, Norway	Norway
Geometry	19	+4	Sweden, Norway, Armenia, Romania	None
Data and Chance	19	+4	Kazakhstan, Ukraine, Armenia, Romania	Armenia
MC Knowing	25	-2	United Arab Emirates, Lebanon	None
MC Applying	20	+3	Norway, Armenia, Romania	None
MC Reasoning	18	+5	Kazakhstan, Sweden, Norway, Armenia, Romania	None
CR Knowing	25	-2	United Arab Emirates, Lebanon	None
CR Applying	21	+2	Armenia, Romania	None
CR Reasoning	19	+4	Ukraine, Norway, Armenia, Romania	None
MC Number	26	-3	United Arab Emirates, Lebanon, Malaysia	None
MC Algebra	23	0		None
MC Geometry	19	+4	Sweden, Norway, Armenia, Romania	None
MC Data and Chance	19	+4	Kazakhstan, Ukraine, Armenia, Romania	Armenia
CR Number	24	-1	United Arab Emirates	None
CR Algebra	19	+4	Slovenia, Sweden, Norway, Romania	None
CR Geometry	21	+2	Sweden, Norway	None
CR Data and Chance	19	+4	Kazakhstan, Ukraine, Armenia, Romania	None

[1] Ranking of Turkey based on overall average percent correct of students is 23.
[2] Dunnett's test results are in Table A7.
[3] None means the mean of Turkey is significantly not different with all countries that it passes or falls down.

Table 30. Dunnett's Test Results for International Average and All Countries

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Korea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Singapore	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Chinese Tai.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006
Hong Kong	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
Japan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
Russian Fed.	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	**0.210**	0.000	0.000	0.029	0.005	0.000	0.006	0.004	0.000	0.005	**0.245**	0.007	0.001	0.001	**0.667**
Israel	0.000	0.000	0.000	0.000	0.000	0.017	0.039	0.029	**0.374**	**0.068**	0.004	0.002	**0.601**	**0.130**	0.037	0.011	**0.318**	0.041	**0.757**	0.038	0.048	**0.063**	**0.804**	**0.593**
Finland	0.000	0.000	0.005	0.001	0.000	**0.250**	0.012	**1.000**	**0.246**	0.001	0.003	0.014	**0.571**	**0.351**	0.012	**0.443**	**0.635**	**1.000**	**0.505**	0.000	0.002	**1.000**	**0.793**	**0.226**
United States	0.000	0.001	**0.058**	0.000	**0.061**	**0.820**	**0.211**	**0.281**	**1.000**	0.009	0.000	**0.114**	**1.000**	**0.179**	**0.526**	**0.799**	**0.766**	**0.100**	**1.000**	0.004	**0.170**	**0.978**	**1.000**	**0.294**
England	0.000	0.004	0.013	**0.075**	0.004	**0.185**	**0.218**	**1.000**	**0.190**	0.000	0.040	**0.093**	**0.811**	**0.993**	0.039	**0.205**	**0.906**	**1.000**	**0.344**	0.000	**0.121**	**1.000**	**0.838**	**0.139**
Hungary	0.002	0.037	0.010	0.001	0.027	**0.366**	**0.239**	**1.000**	**0.074**	**0.070**	0.002	**0.509**	**0.987**	**0.333**	0.049	**0.325**	**0.998**	**1.000**	**0.203**	**0.064**	**0.086**	**1.000**	**0.581**	**0.490**
Australia	0.003	0.012	**0.052**	0.003	0.031	**0.320**	**0.404**	**1.000**	**0.254**	0.003	0.007	**0.198**	**0.708**	**0.428**	**0.184**	**0.491**	**0.996**	**1.000**	**0.327**	0.000	**0.196**	**1.000**	**0.970**	**0.359**
Slovenia	0.038	0.038	**0.331**	0.006	**0.144**	**0.807**	**0.482**	**1.000**	**0.166**	**0.102**	0.001	**0.665**	**0.890**	**0.941**	**0.283**	**0.984**	**0.916**	**1.000**	**0.589**	0.011	**0.429**	**1.000**	**0.470**	**0.971**
Lithuania	**0.054**	**0.065**	**0.337**	0.010	0.035	**1.000**	**0.999**	**1.000**	**0.180**	**0.139**	0.020	**0.199**	**1.000**	**0.552**	**0.213**	**1.000**	**1.000**	**0.999**	**0.325**	**0.060**	**0.748**	**1.000**	**0.837**	**0.829**
Italy	**0.109**	0.018	**0.929**	**0.139**	**0.084**	**0.905**	**0.999**	**1.000**	0.020	**0.558**	**0.122**	**0.079**	**0.508**	**0.992**	**0.778**	**1.000**	**0.795**	**1.000**	**0.061**	**0.245**	**1.000**	**1.000**	**0.469**	**1.000**
New Zealand	**0.705**	**0.990**	**0.696**	**0.421**	**0.627**	**0.988**	**1.000**	**0.942**	**1.000**	**0.050**	**0.482**	**0.999**	**1.000**	**0.999**	**0.661**	**0.926**	**1.000**	**0.769**	**1.000**	0.011	**0.744**	**1.000**	**1.000**	**0.712**
Kazakhstan	**1.000**	**1.000**	**0.916**	**0.998**	**1.000**	**1.000**	**1.000**	**0.986**	**0.895**	**0.931**	**1.000**	**1.000**	**1.000**	**0.925**	**0.941**	**1.000**	**1.000**	**0.922**	**1.000**	**0.356**	**1.000**	**0.989**	**0.269**	**1.000**
Ukraine	**1.000**	**1.000**	**1.000**	**0.997**	**1.000**	**1.000**	**0.999**	**1.000**	**1.000**	**1.000**	**0.983**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.898**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Sweden	**1.000**	**1.000**	**1.000**	**0.999**	**1.000**	**1.000**	**0.976**	**0.127**	**0.760**	**0.289**	**1.000**	**1.000**	**1.000**	**0.995**	**0.990**	**1.000**	**1.000**	**0.064**	**1.000**	**0.273**	**0.305**	**0.914**	**0.626**	**0.822**
Norway	**0.947**	**0.695**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.001	**0.900**	**0.106**	**1.000**	**1.000**	**0.991**	**1.000**	**1.000**	**1.000**	**0.949**	0.000	**1.000**	**0.080**	**0.841**	**0.157**	**0.881**	**0.621**
Armenia	**0.110**	**0.093**	**0.917**	**1.000**	**0.805**	**0.823**	**1.000**	**1.000**	**0.701**	0.003	**0.983**	**0.717**	**0.997**	**1.000**	**1.000**	**0.898**	**1.000**	**0.979**	**0.238**	0.001	**1.000**	**1.000**	**1.000**	**0.290**
Romania	**0.096**	**0.214**	**0.649**	**1.000**	**0.691**	**0.999**	**0.202**	**1.000**	**0.991**	**0.572**	**1.000**	**0.816**	**1.000**	**1.000**	**0.994**	**1.000**	**0.709**	**1.000**	**0.994**	**0.408**	**0.518**	**1.000**	**1.000**	**0.999**
Turkey	**0.743**	**0.919**	**0.971**	**0.616**	**1.000**	**1.000**	**0.276**	**1.000**	**1.000**	**1.000**	**0.680**	**1.000**	**1.000**	**0.999**	**1.000**	**1.000**	**0.594**	**1.000**	**1.000**	**1.000**	**0.558**	**1.000**	**1.000**	**1.000**
U. Arab E.	0.024	**0.143**	**0.162**	**1.000**	**0.081**	**0.432**	**0.475**	**0.995**	**0.108**	**0.961**	**1.000**	**0.323**	**0.739**	**1.000**	**0.324**	**0.667**	**1.000**	**1.000**	**0.465**	**0.773**	**0.501**	**1.000**	**0.391**	**1.000**
Lebanon	0.000	0.000	0.002	**0.900**	0.006	0.034	0.034	**0.563**	**0.157**	0.005	**0.930**	0.015	**0.567**	**1.000**	**0.249**	0.032	**0.936**	**1.000**	**0.104**	0.001	0.026	**0.430**	**0.999**	**0.375**
Malaysia	0.000	0.008	0.001	**0.416**	0.033	**0.119**	**0.241**	0.002	**0.409**	**0.352**	**0.471**	**0.256**	**0.998**	**0.974**	**0.155**	0.043	**1.000**	0.032	**0.985**	**0.303**	**0.176**	0.025	**0.504**	**0.902**
Georgia	0.000	0.001	0.000	**0.308**	0.006	0.041	0.010	**0.381**	0.011	0.049	**0.558**	**0.111**	**0.558**	**0.758**	**0.058**	0.044	**0.106**	**1.000**	**0.231**	**0.117**	**0.119**	**0.344**	**0.073**	**0.244**
Macedonia	0.000	0.000	0.002	0.038	0.002	**0.212**	0.000	**0.824**	**0.138**	0.029	0.039	0.047	**0.669**	**0.778**	0.034	**0.313**	0.041	**0.963**	**0.363**	0.016	0.009	**0.922**	**0.633**	**0.474**
Thailand	0.000	0.000	0.000	0.000	0.000	**0.051**	0.001	0.004	0.004	**0.378**	0.001	0.013	**0.445**	**0.065**	0.022	**0.078**	**0.111**	0.008	**0.067**	**0.410**	0.005	**0.067**	**0.103**	**0.852**
Qatar	0.000	0.000	0.000	0.011	0.000	0.013	0.000	0.007	0.000	0.035	0.017	0.001	**0.183**	**0.428**	0.002	0.035	**0.060**	0.038	0.015	0.024	0.002	**0.109**	0.017	**0.416**
Iran	0.000	0.000	0.000	0.001	0.000	**0.083**	0.000	0.002	**0.144**	0.017	0.004	0.003	**0.707**	**0.108**	0.003	**0.083**	0.015	0.022	**0.308**	0.021	0.000	0.030	**0.741**	**0.238**
Bahrain	0.000	0.000	0.000	0.003	0.000	0.018	0.000	0.001	0.000	**0.214**	0.001	0.000	**0.156**	**0.560**	0.001	**0.061**	0.005	0.006	0.011	0.042	0.000	0.036	0.017	**0.994**
Tunisia	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.005	0.010	0.000	0.000	0.048	**0.108**	0.012	0.013	0.005	0.002	0.027	0.000	0.005	0.006	**0.257**	**0.763**
Palestinian	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.001	0.011	0.002	0.000	0.000	**0.399**	**0.284**	0.001	0.006	0.018	0.010	0.048	0.000	0.000	0.016	**0.347**	**0.242**
Jordan	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.009	0.001	0.005	0.002	0.000	**0.137**	**0.104**	0.000	0.040	0.001	**0.111**	0.010	0.001	0.000	**0.064**	**0.112**	**0.272**
Chile	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.002	**0.340**	0.000	0.003	**0.526**	**0.099**	0.006	0.004	0.003	0.000	**0.187**	**0.603**	0.003	0.003	0.015	**0.614**
Saudi Arabia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	**0.106**	**0.207**	0.000	0.000	0.009	0.000	0.006	0.002	0.000	0.002	0.000	**0.322**
Syrian A. R.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.011	0.000	0.000	0.000	0.002	0.002	0.000	0.000	0.001	0.009	**0.057**
Oman	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.008	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.003	**0.102**
Indonesia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.087**	0.002	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	**0.188**
Morocco	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.003	0.016
Ghana	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004

Bold ones are not significantly different with international average ($p < 0.05$)

obviously fact that, there are three zones on the Table 30, first the top countries; second the countries similar to international average performance (bold ones); and third countries below the international average performance. Generally countries remain also in their zone in different item types. In other words, low performance countries are low and high performance countries are high, in each item type.

CHAPTER 5

DISCUSSION

In this chapter, findings of current research are discussed in the light of literature for better understanding of the study and its findings. This chapter will also include limitations of the study and recommendations for further research.

5.1 The influence of item format on students' performance

The main research question of the study was examining the influence of item format on Turkish students' performance on TIMSS 2011 eighth grade mathematics assessment. This question was examined in terms of three different aspects: MC and CR items, MC and CR items in three different cognitive levels, and MC and CR items in four different content domains. The findings revealed that the students performed significantly better on MC items than CR items. Specifically, a greater percentage of the students answered mathematics MC items correctly than CR items. Similarly, Hastedt (2004) stated that students answered MC items more correctly than CR items. In addition, Ceylan (2013) reported that even the students in high performing schools had difficulties to answer open-ended questions correctly at both fourth and eighth grades.

Çakan (2004), Karakuş (2010), and Kuran and Kanatlı (2009) stated that primary school teachers used mostly MC items as an assessment technique. In addition to this, all national examinations in Turkey (such as TOEG, YGS, and LYS) are formed by only MC items. Hence, the students are familiar to MC items. Therefore, the result of relatively high percentage of students who answered MC

items correctly when it was compared with answering to CR items is in the direction of the expectation.

Additionally, the results showed that the performance of the students were statistically different for the different cognitive level of items. The students' performance on knowing items (M = 44.02) was higher than on applying items (M = 37.55) and reasoning items (M = 29.01). Also, the students' performance on applying items was higher than on reasoning items. Although the main effects of item format and cognitive level of items were significant, the interaction effect was non-significant. Specifically, difference of students' performances on MC and CR items did not change according to cognitive levels.

The highest performance of the students was at knowledge cognitive level which can be seen as a result of the fact that the students met with knowledge level questions frequently and the most of the objectives of the primary mathematics education program were at the knowledge level as well (Coşar, 2010; Güler et al., 2012; Güner, 2015; Kablan et al., 2013). Besides, experiencing applying cognitive level questions mostly in national examinations, in exams prepared by teachers, and in mathematics textbooks were contradicted with the achievement of the students on the TIMSS 2011 eighth grade mathematics assessment on applying items (Delil, 2006; Delil & Tetik, 2015; Dursun & Aydin-Parim, 2014; Güler et al., 2012; Güner, 2015; Incikabi, 2012; Kablan et al., 2013). On the other hand, this result showed a similarity with the percentage of the students correctly answering YGS mathematics applying items (38%) that Dursun and Aydin-Parim (2014) stated in their study.

Similar to this study, Kaleli-Yılmaz and Hancı (2016), Özmen and Karamustafaoğlu (2006) reported that students' performance on reasoning items was very low. Also, Ceylan (2013) stated that eighth grade students' knowing score were

higher than their applying and reasoning scores in low-, medium- and high performing schools in Turkey. The reason for having the lowest performance of the students at reasoning level can be seen that contrary to the TIMSS, national examinations, exams prepared by the mathematics teachers, and textbooks included less questions at this level (Coşar, 2010; Delil & Tetik, 2015; Dursun & Aydin-Parim, 2014; Güner, 2015; Incikabi, 2012; Kablan et al., 2013). Also this failure may be the result of only 12.6% of the objectives of the primary school mathematics program being at the high cognitive level (Kablan et al., 2013).

Thirdly, the results displayed that not all content had the same average of the students' performance. The students' performance on data and chance items (M = 46.13) was significantly higher than their performance on number items (M = 36.03) as well as on algebra items (M = 34.81). However, the students' performances on other content domain items were similar. When Turkish students' performance on TIMSS 1999 and 2007 were examined, it is obviously seen that students' highest performance was still on data and chance regardless of the reform in 2006 (Mullis, Martin, & Foy, 2008; Mullis et.al., 2000). Furthermore, when the Turkish students' performance was compared with international average one, the performance of Turkish students was only above the international average on data and chance. Although the Turkish students' lowest performance was observed on algebra, their performance on number was mostly below the international average. Hence, it could be concluded that the Turkish students' poorest performance was on number. Even though the main effects of item format and content level of items were significant, the interaction effect was non-significant. In other words, difference of students' performances on MC and CR items did not change according to content levels. The

students' MC items performance was higher on all content domains than the students' CR items performance.

The distribution of the contents in national exams, textbooks, objectives of the primary mathematics curriculum, and TIMSS 2011 had differences (Güner, 2015; Incikabi, 2012; Kablan et al., 2013; Uğurel, Moralı, & Kesgin, 2012), and that can cause poor performance, because students have a tendency to ignore the topics that the examinations did not emphasize.

Uzun, Bütüner, and Yiğit (2010) found that students' low achievement on TIMSS 1999 and 2007 could not be explained by attitude, time spent on homework or the education level of parents; on the other hand, they stated that the problems about the application of the new elementary science and mathematics curricula could be a reason. Ünal, Demir, and Kılıç (2011) revealed that professional development affected students' performance in positive way. In addition, the literature showed that cognitive level of the questions that students were exposed to influence their achievement, since students tend to think in the cognitive level that they were used to (Beckmann, 2004; Çepni et al., 2001; Delil & Tetik, 2015; Zhu & Fan, 2006). Consequently, cognitive levels of objectives of the educational program and assessments, as well as teachers' professionality have an importance on students' achievement.

5.2 The change in the ranking of Turkey in different item types

Another important research question was identifying whether Turkey's ranking position changes were significant in different item formats on TIMSS 2011 eighth grade mathematics assessment. In order to examine this question, 42 countries were ranked based on average percent correct of different item types and these rankings

were compared with overall items ranking. The results revealed that the country rankings based on different item types were significantly similar to the ranking based on overall items. More specifically, the rankings based on MC items (21st rank) and CR items (23rd rank) were similar to the ranking based on overall items (23rd rank) differently from the expectation.

Further, the performances of Turkey on different item types were compared with their international averages. The results showed that the performance of Turkey in each item type was not significantly different from its international average. The performance of Turkey was only above the international average on data and chance, MC data and chance, and CR data and chance items. Besides, Turkish students' performance was mostly below the international average on knowing, number, MC number and CR number items. Whether the performance of Turkey below or above the international averages, the difference between them was not significant. That is to say, the performance of Turkey below or above the international average in different item type does not indicate that the performance of Turkey is better or worse than its international average. In addition, when the international average and the average performance of other countries were compared, it was concluded that low performance countries were low and high performance countries were high in each item type.

Moreover, the rank of Turkey was highest (18th rank) on reasoning and MC reasoning items. Turkey passed five countries in each ranking according to overall item ranking, but there were no significant difference between the performances of the students from these countries. On the other side, since Turkish students' average performance on reasoning items was 29.01%, the highest ranking does not mean that the students' performance was high on reasoning items. The lowest rank of Turkey

(26th rank) was on the ranking of number and MC number items. Turkey fell down United Arab Emirates, Lebanon and Malaysia, yet there were no significant differences between the performances of the students.

The studies in the literature have similarities to this research. Hastedt and Sibberns (2005) compared ranking based on MC and CR items by using TIMSS 1995 and TIMSS 1999 data, while Nixon and Barth (2014) compared students' performance from five countries on cognitive domains by using some TIMSS items, and Beaton (1998) compared fourth grade students' performance from 24 countries by TIMSS mathematics content sub-areas. Adams et al. (2010) studied on an analysis of PISA 2006 preferred items ranking. Hastedt and Sibberns (2005) reported that both TIMSS 1995 and TIMSS 1999 mathematics rankings based on MC items, CR items and total scale were very similar. According to Nixon and Barth's (2014) study, changes in the rankings did not show difference in students' performance when items grouped by domain. Adams et al. (2010) found that the ranking positions of the several countries could be altered systematically but the sampling error could explain the shifts in ranking positions of these countries. On the other hand, systematic shifts were not observed for most of the countries. Beaton (1998) showed that performances in the different sub-areas were highly correlated. If the students were successful in one sub-area, they were more likely to do well in other sub-areas. In other words, it was reported that the top-scoring countries are success in all sub-areas, the middle-scorers do middling well, and low-scorers perform less well. This study extended these researches. It enclosed all items and all countries participated in TIMSS 2011 eighth grade mathematics assessment. Also, it examined the ranking chances based on item format, cognitive and content domains of items. However, the findings were very similar.

All these findings showed that, Turkey's position is in the middle whatever the item type. As a result, our expectation on Turkey's significant position chances favor on MC items was not found. Difference between the means of performance of the students on MC and CR items was between 11.44 and 19.89 among all 42 countries. In Turkey, this difference is 16.58 and very close to the difference of international averages (16.94). This result displayed that all countries had higher performance on MC items.

5.3 Limitations of the study

This study aimed to compare the performance of the students based on all item types. However, number of items on the research was not adequate in order to analyze item format, cognitive and content domains of the items together (e.g. MC knowing number items, CR applying algebra items etc.) (See Table 8).

5.4 Recommendations and suggestions for further research

This is the first study in Turkey and especially important to give information about students' performance on different item types. Educators and policy makers can see the weaknesses of the students and what areas need to be improved on mathematics education.

The cognitive level of the questions asked to the students makes an effect on their perfromance as they affect their tendency to think (Beckmann, 2004; Çepni et.al., 2001; Delil & Tetik, 2015; Zhu & Fan, 2006). For this reason, in our education system, where only 12.6% of primary school mathematics program objectives are at reasoning level (Kablan et.al., 2013), it is possible to increase students' thinking skills at higher cognitive level by the changes in the objectives of the mathematics

curriculum or by in-class activities. Also, the weight of higher cognitive level of questions on the teachers' exams, on the textbooks or asked by teachers should be increased.

On the other side, the unexpected performance of the students on MC items should be examined in detail, since the result contradicted with the students' familiarity on MC items. Besides, the reasons of the students' poor performance especially on number and algebra should be examined accurately. After these studies, the educational program can be improved.

Further research may also examine influence of item type on science performance of the students. Since mathematics and science are related with each other, the results can be evaluated together. This study focused on eighth grade students' mathematics performance, another research can examine fourth grade students' achievement. Additionally, it would be interesting to conduct similar study on trends (spans across multiple years of data). Hence, an improvement in a particular domain can be seen over time and it can be compared with other nations.

APPENDIX A

THE NUMBER OF STUDENTS AND SCHOOLS

Table A1. The Number of Students and Schools Attended TIMSS 2011 Eighth Grade Mathematics Assessment

Name of the Country	Number of Schools	Number of Students
Armenia	153	5,846
Australia	277	7,556
Bahrain	95	4,640
Chile	193	5,835
Chinese Taipei	150	5,042
England	118	3,842
Finland	145	4,266
Georgia	172	4,563
Ghana	161	7,323
Hong Kong SAR	117	4,015
Hungary	146	5,178
Indonesia	153	5,795
Iran, Islamic Rep. of	238	6,029
Israel	151	4,699
Italy	197	3,979
Japan	138	4,414
Jordan	230	7,694
Kazakhstan	147	4,390
Korea, Rep. of	150	5,166
Lebanon	147	3,974
Lithuania	141	4,747
Macedonia, Rep. of	150	4,062
Malaysia	180	5,733
Morocco	279	8,986
New Zealand	158	5,336
Norway	134	3,862
Oman	323	9,542
Palestinian Natl. Auth.	201	7,812
Qatar	109	4,422
Romania	147	5,523
Russian Federation	210	4,893
Saudi Arabia	153	4,344
Singapore	165	5,927
Slovenia	186	4,415
Sweden	153	5,573
Syrian Arab Rep.	148	4,413
Thailand	172	6,124
Tunisia	207	5,128
Turkey	239	6,928
Ukraine	148	3,378
United Arab Emirates	458	14,089
United States	501	10,477
Total	7,840	239,960

(Joncas, 2013)

APPENDIX B

MATHEMATICS CONTENT AND COGNITIVE DOMAINS

Table A2. Mathematics Content Domains, Topic Areas, and Their Objectives

Content Domains	Topic Areas	Objectives/Behaviours
Number	Whole numbers	1. Demonstrate understanding of the principles of whole numbers and operations with them (e.g., knowledge of the four operations, place value, commutativity, associativity and distributivity). 2. Find and use multiples or factors of numbers, identify prime numbers, and evaluate powers of numbers and square roots of perfect squares to 144. 3. Solve problems by computing, estimating, or approximating with whole numbers.
	Fractions and decimals	1. Compare and order fractions; recognize and write equivalent fractions. 2. Demonstrate understanding of place value for finite decimals (e.g., by comparing or ordering them). 3. Represent fractions and decimals and operations with fractions and decimals using models (e.g., number lines); identify and use such representations. 4. Convert between fractions and decimals. 5. Compute with fractions and decimals and solve problems involving them.
	Integers	1. Represent, compare, order, and compute with integers and solve problems using them.
	Ratio, proportion and percent	1. Identify and find equivalent ratios; model a given situation by using a ratio and divide a quantity in a given ratio. 2. Convert between percents and fractions or decimals. 3. Solve problems involving percents and proportions.
Algebra	Patterns	1. Extend well-defined numeric, algebraic, and geometric patterns or sequences using numbers, words, symbols, or diagrams; find missing terms. 2. Generalize pattern relationships in a sequence, or between adjacent terms, or between the sequence number of the term and the term, using numbers, words, or algebraic expressions.
	Algebraic expressions	1. Find sums, products, and powers of expressions containing variables. 2. Evaluate expressions for given numeric values of the variable(s). 3. Simplify or compare algebraic expressions to determine if they are equal. 4. Model situations using expressions.
Algebra	Equations/ formulas and functions	1. Evaluate equations/formulas given values of the variables. 2. Indicate whether a value (or values) satisfies a given equation/formula. 3. Solve linear equations and linear inequalities, and simultaneous (two variables) linear equations. 4. Recognize and write equations, inequalities, simultaneous equations, or functions that model given situations. 5. Recognize and generate representations of functions in the form of tables, graphs, or words. 6. Solve problems using equations/formulas and functions.
Geometry	Geometric shapes	1. Identify different types of angles and know and use the relationships between angles on lines and in geometric figures. 2. Recognize geometric properties of common two-and-three-dimensional shapes, including line and rotational symmetry. 3. Identify congruent triangles and quadrilaterals and their corresponding measures; identify similar triangles and recall and use their properties. 4. Recognize relationships between three-dimensional shapes and their two-dimensional representations (e.g., nets or two-dimensional views of three-dimensional objects). 5. Apply geometric properties, including the Pythagorean Theorem, to solve problems.

Content Domains	Topic Areas	Objectives/Behaviours
Geometry	Geometric measurements	1. Draw given angles and lines; measures and estimate the size of given angles, line segments, perimeters, areas, and volumes. 2. Select and use appropriate measurement formulas for perimeters, circumferences, areas, surface areas, and volumes; find measures of compound areas. 3. Locate points in the Cartesian plane, and solve problems including such points. 4. Recognize and use geometric transformations (translation, reflection, and rotation) of two-dimensional shapes.
	Location and movement	1. Read scales and data from tables, pictographs, bar graphs, pie charts, and line graphs. 2. Organize and display data using tables, pictographs, bar graphs, pie charts, and line graphs. 3. Compare and match different representations of the same data.
Data and Chance	Data organization and representation	1. Identify, calculate and compare characteristics of data sets, including mean, median, mode, range and shape of distribution (in general terms). 2. Use and interpret data sets to answer questions and solve problems (e.g., make inferences, draw conclusions, and estimate values between and beyond given data points). 3. Recognize and describe approaches to organizing and displaying data that could lead to misinterpretation (e.g., inappropriate grouping a misleading or distorted scales).
	Data interpretation	1. Judge the chance of an outcome as certain, more likely, equally likely, less likely, or impossible. 2. Use data to estimate the chances of future outcomes; use the chances of a particular outcome to solve problems; determine the chances of possible outcomes.
	Chance	1. Judge the chance of an outcome as certain, more likely, equally likely, less likely, or impossible. 2. Use data to estimate the chances of future outcomes; use the chances of a particular outcome to solve problems; determine the chances of possible outcomes.

(Mullis et al., 2009)

Table A3. Item Labels in each Content and Cognitive Domain

Content Domain	Topic Areas	Cognitive Domain	Item Label
Number	Whole numbers	Knowing	36 as a product of prime factors Approximate the sum Equivalent numeric expression What is the value of cube of 3 Which is a prime number
			Packing eggs into boxes Height of a stack of stools Express 256 × 4096 as power of 4 How many kilocalories used How many minutes of soccer played Answer to Robert's problem Cost of phone for 1 year Pay after first year Time when Pat finishes last lap
		Reasoning	Which statement about x is true Numbers to get greatest results Complete TV game show table Which plan is less expensive Multiples of 3
	Fractions and decimals	Knowing	Best estimate of (7.21×3.86)/10.09 Select the decimal equal to 3/5 Add 42.65 to 5.748 Which number sentence is true Equivalent fraction for 0.125 Decimal closest in size to 3/4. Compare size of decimal numbers 4/100 plus 3/1000 What is K on a number line Write 3 5/6 in decimal form Fraction wanting to go on a trip Find the number Complete the fraction sequence Fraction of diagram shaded Write as decimal number Octagon with equivalent shading Convert decimal to a fraction.
		Applying	Method for subtracting fractions Equivalent expression Ann and Jenny divide 560 zeds Length of the original pipe Which is larger 7/12 or 2/3 How far did john jump
		Reasoning	Location of N on number line Who spent more for shoes
	Integers	Knowing	Solve given numeric expression How much hotter is city A than B
		Reasoning	Next line in the pattern Rule to get terms in the pattern
	Ratio, proportion and percent	Knowing	Complete the missing boxes Fraction by which price reduced by Present of games the team lose Complete the table Ratio equivalent to 1:4

Content Domain	Topic Areas	Cognitive Domain	Item Label
Number	Ratio, proportion and percent	Applying	The percentage of caps for sale Milk needed in cake recipe Percentage of money John saves Number of papers in a 28mm stack Length of shorter string piece Which statement is true Percentage of laps finished
		Reasoning	Who paid less for a hockey stick
Algebra	Patterns	Knowing	Next term in the pattern
		Applying	The shadow lengths of four bushes
		Reasoning	Red&BlackTiles_Complete table Red&BlackTiles_Shape with 64 tiles Red&BlackTiles_Shape with 49 tiles Red&BlackTiles_Shape with 44 tiles Value of x in the pattern Term number 100 in the pattern Term number n in the pattern Missing term in the sequence Area of the 5th square Area of the nth square Circles for patterns 4 & 30 Procedure for finding the number Rule for finding number of circles Number of matches for figure 10 Rule for number of matches
	Algebraic expressions	Knowing	If t is a number between 6 and 9 Simplify the expression Value of the given expression Which represents 2x plus 3x Cost in zeds for taxi trip of n km Expression to equivalent to 4(3+x) Simplify 3x/8 plus x/4 plus x/2 What xy plus 1 mean m boys and n girls in a parade Equivalent algebraic expression Value of fractional expression What is a+b What is a-b Value of the expression What does (ab +1)2 equal Identify an equivalent expression Expression for one-fifth of x Expression with exponents of y
		Applying	Area of garden's shaded portion Expression for the shaded area What is the area of the rectangle Sum of 3 consecutive whole numbers Expression for area of rectangle
		Reasoning	Red&BlackTiles_Figure n

Content Domain	Topic Areas	Cognitive Domain	Item Label
Algebra	Equations/ formulas and functions	Knowing	Value of y in an expression What is the value P Solve the inequality Value of y when x = 25 Find the value of y when t is 9 What is the value of x and y Convert Celsius into Fahrenheit What is the value of x Which equation is satisfied What is the value of n Length of longest side of triangle What is value of 3x+3y Pair of numbers that satisfy eqn
		Applying	Length of the longest wood piece What is the value of 2a + 2b + 4 Equation that satisfy number pairs Formula for the new rectangle Which slope has a greater value Write an equation to find x Step in solving the equation Set up system of equations Relationship between x and y Formula for distance traveled Solve for simultaneous equations Formula for K the cost of trip Value of x and y Temperature at mountain top Graph of 3y = 2x + 6
		Reasoning	Jo has 3 metal blocks to weigh
Geometry	Geometric shapes	Knowing	View of shape directly from above Which shape has a line of symmetry Shape of cutout figure Statement about Lines L and M
		Applying	Draw an isosceles triangle Value of angle x in figure Degrees minute hand of clock turns What is the value of angle a What is the value of x Length of the side of the triangle Value of angle y in a triangle Shape with AB as line of symmetry Lines of symmetry for pentagon Length of line segment AB How many cubes were left Value of angle x outside triangle
		Reasoning	Interior angles of pentagon What is the size of angle B Value of angle b Height of the building Figure that can be folded as box Distance between the midpoints Why PQR is a right angle triangle Value of angle x in figure Conditions for similar triangles What is the perimeter of the figure Which 2 shapes fit together Area of the triangle ABC

Content Domain	Topic Areas	Cognitive Domain	Item Label
Geometry	Geometric shapes	Reasoning	Interior angles of pentagon What is the size of angle B Value of angle b Height of the building Figure that can be folded as box Distance between the midpoints Why PQR is a right angle triangle Value of angle x in figure Conditions for similar triangles What is the perimeter of the figure Which 2 shapes fit together Area of the triangle ABC
	Geometric measurements	Applying	Calculate the area of a square Length of the rectangular box Area of a square is 144 square cm Shape made up of same size cubes Measure of angle BOC Area of the shaded region in figure Draw an angle ABC of 135 degrees Surface area of the prism
		Reasoning	Number of books to fill the box Number of cubes that fit in a box
	Location and movement	Knowing	Half-turn around point O Figure 1 transformed to 2 and 3
		Applying	Mark the points B and C on graph
		Reasoning	Point equidistant from P and Q Rotation and reflection
Data and Chance	Data organization and representation	Knowing	Car production graph/time cars made Number of goals of 4 soccer teams Pizzas sold by a shop Speed on the speedometer Girls playing video games
		Applying	Complete and label this pie chart Students that chose History Make a pie chart with labels Store with lowest price per pair Representing data on a pie chart Complete the bar chart Liv's smartphone use
	Data interpretation	Knowing	Mean number of staff members Median number of staff members Estimate the temperature Difference in high and low temp Difference in percentages
		Applying	List with specified mean and range Change in mean and median Car production graph/avg by hour Car production graph/identify time Is Jesse's bar graph correct Speed time graph Raspberry as favorite flavor Smartphone use listening to music

Content Domain	Topic Areas	Cognitive Domain	Item Label
Data and Chance	Data interpretation	Reasoning	Sales of two types of soft drink Age structures of country X and Y Problem of taking care of elderly Long jump competition Sports survey bar graph mistake Explain why Pat is correct Agree/disagree with the salesman Overall average of 9 for Ahmed
	Chance	Knowing	How likely to get pink candy How likely student voted for Pat How likely it will rain
		Applying	Chance of getting a button Number of regular size bottles Number of times spinner in red area Blue color on the spinners Red color on the spinners Number of balls in a bag
		Reasoning	Probability that the marble is red

(IEA, 2013b)

APPENDIX C

THE NUMBER OF STUDENTS ANSWERED THE ITEM

Table A4. Items and the Number of Students Answered It

Item ID	Content Domain	Cognitive Domain	Item Format	Number of Studs.
M032166	Number	Knowing	Multiple Choice	997
M032721	Data and Chance	Reasoning	Multiple Choice	997
M032757	Algebra	Reasoning	Constructed Response	997
M032760A	Algebra	Reasoning	Constructed Response	997
M032760B	Algebra	Reasoning	Constructed Response	997
M032760C	Algebra	Reasoning	Constructed Response	997
M032761	Algebra	Reasoning	Constructed Response	997
M032692	Geometry	Reasoning	Constructed Response	997
M032626	Number	Knowing	Multiple Choice	997
M032595	Number	Applying	Multiple Choice	997
M032673	Algebra	Knowing	Multiple Choice	997
M052216	Number	Knowing	Multiple Choice	990
M052231	Number	Knowing	Constructed Response	990
M052061	Number	Applying	Constructed Response	990
M052228	Number	Applying	Multiple Choice	990
M052214	Number	Knowing	Multiple Choice	990
M052173	Algebra	Applying	Multiple Choice	990
M052302	Algebra	Knowing	Multiple Choice	990
M052002	Algebra	Applying	Constructed Response	990
M052362	Geometry	Reasoning	Constructed Response	990
M052408	Geometry	Reasoning	Constructed Response	990
M052084	Geometry	Applying	Multiple Choice	990
M052206	Geometry	Reasoning	Constructed Response	990
M052429	Data and Chance	Reasoning	Multiple Choice	990
M052503A	Data and Chance	Reasoning	Constructed Response	990
M052503B	Data and Chance	Reasoning	Constructed Response	990
M042032	Number	Knowing	Multiple Choice	985
M042031	Number	Applying	Multiple Choice	985
M042186	Number	Reasoning	Constructed Response	985
M042059	Number	Knowing	Constructed Response	985
M042236	Algebra	Knowing	Multiple Choice	985
M042226	Algebra	Knowing	Constructed Response	985
M042103	Algebra	Knowing	Constructed Response	984
M042086	Algebra	Applying	Constructed Response	984
M042228	Algebra	Reasoning	Constructed Response	984
M042245	Algebra	Applying	Multiple Choice	984
M042270	Geometry	Applying	Constructed Response	985
M042201	Geometry	Applying	Constructed Response	985
M042152	Geometry	Knowing	Multiple Choice	985
M042269	Data and Chance	Reasoning	Multiple Choice	985
M042179	Data and Chance	Applying	Multiple Choice	985
M042177	Data and Chance	Applying	Multiple Choice	984
M042207	Data and Chance	Applying	Constructed Response	985
M052209	Number	Knowing	Multiple Choice	993
M052142	Number	Applying	Multiple Choice	993
M052006	Number	Reasoning	Multiple Choice	993
M052035	Number	Knowing	Constructed Response	993
M052016	Number	Applying	Constructed Response	993
M052064	Algebra	Knowing	Multiple Choice	993
M052126	Algebra	Applying	Constructed Response	993
M052103	Algebra	Knowing	Multiple Choice	993
M052066	Algebra	Applying	Multiple Choice	993

Item ID	Content Domain	Cognitive Domain	Item Format	Number of Studs.
M052041	Geometry	Reasoning	Constructed Response	993
M052057	Geometry	Reasoning	Multiple Choice	993
M052417	Geometry	Applying	Constructed Response	993
M052501	Data and Chance	Reasoning	Constructed Response	993
M052410	Data and Chance	Applying	Multiple Choice	992
M052170	Data and Chance	Applying	Multiple Choice	993
M032094	Number	Knowing	Multiple Choice	986
M032662	Number	Reasoning	Multiple Choice	986
M032064	Number	Applying	Constructed Response	985
M032419	Algebra	Knowing	Multiple Choice	986
M032477	Algebra	Knowing	Multiple Choice	986
M032538	Algebra	Knowing	Constructed Response	986
M032324	Geometry	Reasoning	Multiple Choice	986
M032116	Geometry	Applying	Multiple Choice	986
M032100	Geometry	Applying	Multiple Choice	986
M032402	Geometry	Reasoning	Multiple Choice	986
M032734	Geometry	Knowing	Constructed Response	986
M032397	Geometry	Knowing	Multiple Choice	986
M032695	Data and Chance	Applying	Constructed Response	986
M032132	Data and Chance	Knowing	Multiple Choice	986
M042041	Number	Applying	Multiple Choice	982
M042024	Number	Knowing	Multiple Choice	982
M042016	Number	Applying	Multiple Choice	981
M042002	Number	Reasoning	Constructed Response	981
M042198A	Algebra	Knowing	Constructed Response	981
M042198B	Algebra	Reasoning	Constructed Response	981
M042198C	Algebra	Reasoning	Constructed Response	981
M042077	Algebra	Knowing	Multiple Choice	981
M042235	Algebra	Knowing	Multiple Choice	982
M042067	Algebra	Applying	Multiple Choice	982
M042150	Geometry	Knowing	Multiple Choice	982
M042300A	Geometry	Applying	Constructed Response	981
M042300B	Geometry	Applying	Constructed Response	981
M042260	Data and Chance	Knowing	Multiple Choice	981
M042169A	Data and Chance	Knowing	Constructed Response	982
M042169B	Data and Chance	Knowing	Constructed Response	982
M042169C	Data and Chance	Applying	Constructed Response	982
M032352	Algebra	Applying	Multiple Choice	986
M032725	Number	Knowing	Constructed Response	986
M032683	Algebra	Knowing	Constructed Response	986
M032738	Algebra	Knowing	Multiple Choice	986
M032295	Algebra	Knowing	Multiple Choice	986
M032331	Geometry	Applying	Multiple Choice	986
M032623	Geometry	Applying	Multiple Choice	986
M032679	Geometry	Knowing	Multiple Choice	986
M032047	Algebra	Applying	Multiple Choice	986
M032398	Geometry	Reasoning	Multiple Choice	986
M032507	Data and Chance	Applying	Multiple Choice	986
M032424	Algebra	Reasoning	Multiple Choice	986
M032681A	Data and Chance	Knowing	Constructed Response	986
M032681B	Data and Chance	Applying	Constructed Response	986
M032681C	Data and Chance	Applying	Constructed Response	985
M052413	Number	Knowing	Multiple Choice	988
M052134	Number	Knowing	Multiple Choice	988
M052078	Number	Applying	Multiple Choice	988
M052034	Number	Knowing	Multiple Choice	988
M052174A	Number	Applying	Constructed Response	988
M052174B	Number	Applying	Constructed Response	988
M052130	Algebra	Knowing	Multiple Choice	988

Item ID	Content Domain	Cognitive Domain	Item Format	Number of Studs.
M052073	Algebra	Knowing	Multiple Choice	988
M052110	Algebra	Knowing	Constructed Response	986
M052105	Algebra	Applying	Constructed Response	986
M052407	Geometry	Applying	Multiple Choice	988
M052036	Geometry	Applying	Constructed Response	985
M052502	Data and Chance	Applying	Constructed Response	985
M052117	Data and Chance	Applying	Constructed Response	985
M052426	Data and Chance	Knowing	Multiple Choice	988
M042183	Number	Knowing	Multiple Choice	996
M042060	Number	Knowing	Multiple Choice	996
M042019	Number	Knowing	Constructed Response	993
M042023	Number	Applying	Constructed Response	993
M042197	Number	Reasoning	Constructed Response	993
M042234	Algebra	Knowing	Multiple Choice	996
M042066	Algebra	Reasoning	Constructed Response	993
M042243	Algebra	Knowing	Multiple Choice	996
M042248	Algebra	Knowing	Constructed Response	993
M042229A	Algebra	Applying	Constructed Response	993
M042229B	Algebra	Knowing	Constructed Response	993
M042080A	Algebra	Knowing	Constructed Response	993
M042080B	Algebra	Knowing	Constructed Response	993
M042120	Geometry	Knowing	Multiple Choice	996
M042203	Geometry	Applying	Multiple Choice	996
M042264	Geometry	Reasoning	Constructed Response	993
M042255	Data and Chance	Applying	Multiple Choice	996
M042224	Data and Chance	Knowing	Constructed Response	993
M052017	Number	Knowing	Multiple Choice	990
M052217	Number	Reasoning	Constructed Response	990
M052021	Number	Reasoning	Constructed Response	988
M052095	Number	Knowing	Constructed Response	990
M052094	Number	Reasoning	Constructed Response	990
M052131	Algebra	Applying	Multiple Choice	990
M052090	Algebra	Applying	Multiple Choice	990
M052121A	Algebra	Reasoning	Multiple Choice	990
M052121B	Algebra	Reasoning	Constructed Response	990
M052042	Geometry	Applying	Constructed Response	990
M052047	Geometry	Applying	Constructed Response	990
M052044	Geometry	Reasoning	Multiple Choice	989
M052422A	Data and Chance	Applying	Multiple Choice	990
M052422B	Data and Chance	Applying	Multiple Choice	990
M052505	Data and Chance	Knowing	Multiple Choice	990
M042015	Number	Knowing	Multiple Choice	984
M042196	Data and Chance	Knowing	Multiple Choice	984
M042194	Number	Knowing	Constructed Response	984
M042114A	Number	Knowing	Constructed Response	984
M042114B	Number	Applying	Constructed Response	984
M042112	Algebra	Applying	Multiple Choice	984
M042109	Algebra	Applying	Multiple Choice	984
M042050	Algebra	Knowing	Constructed Response	984
M042074A	Algebra	Reasoning	Constructed Response	984
M042074B	Algebra	Reasoning	Constructed Response	984
M042074C	Algebra	Reasoning	Constructed Response	984
M042151	Geometry	Applying	Constructed Response	984
M042132	Geometry	Reasoning	Multiple Choice	984
M042257	Geometry	Reasoning	Multiple Choice	983
M042158	Data and Chance	Knowing	Multiple Choice	984
M042252	Data and Chance	Applying	Multiple Choice	984
M042261	Data and Chance	Knowing	Multiple Choice	984
M052079	Number	Applying	Multiple Choice	982

Item ID	Content Domain	Cognitive Domain	Item Format	Number of Studs.
M052204	Number	Knowing	Multiple Choice	982
M052364	Number	Applying	Constructed Response	982
M052215	Number	Knowing	Constructed Response	982
M052147	Number	Applying	Multiple Choice	981
M052067	Algebra	Knowing	Multiple Choice	982
M052068	Algebra	Knowing	Multiple Choice	982
M052087	Algebra	Applying	Constructed Response	982
M052048	Geometry	Applying	Constructed Response	982
M052039	Geometry	Applying	Constructed Response	982
M052208	Geometry	Reasoning	Multiple Choice	982
M052419A	Data and Chance	Knowing	Multiple Choice	982
M052419B	Data and Chance	Knowing	Multiple Choice	982
M052115	Data and Chance	Applying	Multiple Choice	982
M052421	Data and Chance	Reasoning	Constructed Response	982
M042182	Number	Applying	Multiple Choice	990
M042081	Number	Knowing	Constructed Response	990
M042049	Algebra	Knowing	Multiple Choice	990
M042052	Number	Knowing	Multiple Choice	990
M042076	Algebra	Knowing	Multiple Choice	990
M042302A	Number	Applying	Constructed Response	989
M042302B	Number	Applying	Constructed Response	989
M042302C	Number	Reasoning	Constructed Response	989
M042100	Algebra	Knowing	Multiple Choice	989
M042202	Algebra	Applying	Multiple Choice	989
M042240	Algebra	Applying	Multiple Choice	989
M042093	Algebra	Applying	Constructed Response	989
M042271	Geometry	Applying	Multiple Choice	989
M042268	Geometry	Reasoning	Multiple Choice	989
M042159	Data and Chance	Applying	Constructed Response	989
M042164	Data and Chance	Reasoning	Constructed Response	989
M042167	Data and Chance	Reasoning	Constructed Response	989
M052024	Number	Knowing	Multiple Choice	998
M052058A	Number	Applying	Constructed Response	999
M052058B	Number	Applying	Constructed Response	999
M052125	Number	Reasoning	Multiple Choice	999
M052229	Number	Knowing	Constructed Response	999
M052063	Algebra	Applying	Multiple Choice	999
M052072	Algebra	Knowing	Multiple Choice	999
M052146A	Algebra	Reasoning	Constructed Response	999
M052146B	Algebra	Reasoning	Constructed Response	999
M052092	Algebra	Applying	Multiple Choice	999
M052046	Geometry	Reasoning	Multiple Choice	999
M052083	Geometry	Applying	Multiple Choice	999
M052082	Geometry	Applying	Multiple Choice	999
M052161	Data and Chance	Applying	Multiple Choice	999
M052418A	Data and Chance	Applying	Multiple Choice	999
M052418B	Data and Chance	Applying	Multiple Choice	999

(IEA, 2013b)

APPENDIX D

ANALYSIS TABLES

Table A5. Average Percentages Correct at Eighth Grade for 42 Countries across Different Item Type in TIMSS

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Armenia	38.90	45.80	30.68	50.18	36.46	25.55	44.11	41.84	35.41	30.23	50.73	41.15	35.97	42.63	20.89	20.03	50.75	50.70	35.28	37.56	32.66	26.41	35.59	20.05
Australia	47.46	54.75	38.77	56.68	46.39	35.01	51.93	38.33	45.79	57.64	63.45	50.72	46.03	50.21	41.03	29.18	55.11	45.61	49.66	67.01	44.09	27.05	40.41	44.63
Bahrain	29.51	37.29	20.24	37.34	27.60	20.59	29.98	26.86	27.22	35.46	43.19	34.44	30.09	29.72	19.14	15.56	38.24	35.42	31.38	41.72	18.57	14.90	21.44	26.77
Chile	28.55	36.89	18.69	33.81	28.91	19.87	29.89	22.30	29.02	36.34	41.91	35.53	32.73	25.28	20.89	13.05	37.56	30.68	34.83	46.08	20.79	12.34	21.27	22.82
Chinese Taipei	70.89	76.71	63.96	77.17	71.94	59.54	72.05	71.37	73.04	66.34	80.36	74.78	70.18	70.11	68.42	53.91	77.01	76.18	75.00	75.29	64.51	63.19	70.32	53.91
England	48.23	55.30	39.81	56.80	47.79	35.77	52.44	38.99	46.18	59.36	61.63	51.54	45.52	45.14	43.16	30.61	56.24	48.06	49.56	67.96	44.89	28.82	41.49	47.43
Finland	49.01	56.32	40.31	57.60	49.28	35.37	55.57	39.06	45.83	59.10	64.16	53.18	46.69	50.80	44.46	29.37	57.52	48.77	48.78	68.46	49.68	25.56	41.74	46.09
Georgia	33.47	42.47	22.74	43.16	31.77	21.33	35.69	33.59	30.37	33.21	48.98	38.57	32.89	30.95	23.37	15.21	41.66	44.21	35.23	43.07	25.28	18.35	23.62	19.51
Ghana	19.06	27.89	8.52	25.08	17.68	12.05	19.28	18.58	17.77	20.80	30.41	25.76	24.53	14.53	7.69	5.44	29.11	27.78	23.79	27.98	6.61	6.66	9.40	10.83
Hong Kong	66.98	73.32	59.42	75.56	67.27	53.32	70.68	63.02	69.36	65.80	80.02	69.65	63.14	67.20	64.32	48.12	75.39	69.74	69.43	73.92	62.26	50.98	69.26	54.52
Hungary	48.10	54.78	40.14	58.46	46.49	34.81	52.56	41.14	47.29	53.91	64.53	49.42	44.06	50.94	42.87	29.91	55.39	47.61	50.48	61.40	45.41	28.87	42.86	43.50
Indonesia	24.45	33.15	14.07	30.76	23.33	16.55	24.68	21.80	24.70	28.17	38.00	29.92	29.10	18.89	15.18	9.90	34.04	31.11	30.25	34.95	12.45	9.73	16.99	18.76
Iran	30.01	38.55	19.82	36.52	28.74	22.04	30.21	26.87	32.90	31.92	43.94	35.66	33.62	25.45	20.20	15.91	39.29	36.70	35.73	40.91	17.97	14.69	28.97	19.44
Israel	50.60	57.56	42.29	59.85	49.48	38.20	54.68	48.29	45.20	53.95	63.89	54.54	46.54	53.22	43.21	33.78	59.04	56.53	47.74	62.05	46.23	36.01	41.68	42.71
Italy	45.75	54.68	35.10	54.31	45.57	32.86	48.84	38.64	48.54	50.14	60.98	51.70	47.02	45.58	38.00	25.37	56.83	48.28	52.17	59.40	36.59	24.92	43.49	37.29
Japan	63.23	68.84	56.55	69.26	64.09	52.57	62.94	59.50	66.32	66.64	75.00	67.40	58.99	63.08	60.00	49.17	66.68	67.01	65.11	74.86	56.83	49.11	68.01	55.22
Jordan	29.04	37.53	18.92	37.05	26.91	20.21	28.40	29.02	28.41	30.62	43.48	33.69	29.86	25.37	18.52	15.10	36.98	38.53	31.33	38.37	16.22	15.64	24.35	19.85
Kazakhstan	42.82	48.40	36.16	52.58	41.18	30.46	44.38	43.32	43.31	39.29	54.36	44.45	39.13	46.83	37.15	25.87	49.09	51.21	42.18	44.92	34.64	29.74	44.88	31.47
Korea	72.85	78.14	66.55	79.84	73.31	61.36	77.29	69.99	71.50	72.56	83.44	77.67	65.23	72.99	667.91	59.31	80.70	75.71	70.70	82.83	71.22	60.49	72.62	58.31
Lebanon	34.74	43.19	24.68	46.75	31.76	21.14	38.02	35.51	33.01	30.57	50.29	36.82	32.94	34.68	25.49	14.89	46.46	45.18	34.03	37.53	23.18	18.81	31.59	20.90
Lithuania	46.28	54.03	37.05	56.22	46.28	31.00	48.68	40.16	46.25	52.89	62.66	50.72	42.16	49.38	40.79	25.10	51.76	49.98	49.67	61.46	41.03	25.60	41.49	40.98
Macedonia	32.79	40.74	23.40	41.04	31.03	23.15	32.86	32.88	32.85	32.49	46.20	37.74	33.44	31.59	22.72	17.70	40.42	43.22	36.04	40.54	22.01	21.28	28.41	21.76
Malaysia	34.23	43.20	23.54	43.30	32.91	22.44	40.04	27.74	34.37	36.42	48.69	39.53	36.13	32.93	24.72	15.19	48.26	37.08	39.17	44.61	25.92	14.49	27.71	25.03
Morocco	22.69	30.65	13.21	28.25	22.65	14.23	24.14	19.63	25.05	23.27	33.85	28.67	26.48	17.73	15.19	7.74	31.35	29.25	29.28	30.46	13.01	8.04	19.18	13.28
New Zealand	43.67	50.40	35.65	51.53	43.26	32.23	47.33	34.87	42.09	54.38	59.17	47.13	41.18	44.88	38.48	27.50	50.22	41.96	45.50	63.39	41.05	24.86	37.37	41.86
Norway	39.00	45.16	31.73	46.22	39.37	27.31	45.99	25.69	36.14	53.32	52.26	43.92	35.71	43.58	33.87	22.86	46.57	33.09	40.72	61.10	40.58	16.95	29.77	42.51
Oman	24.53	32.99	14.45	31.40	22.91	16.64	24.15	23.11	25.22	26.72	37.42	30.03	27.77	21.04	14.09	10.75	33.97	31.81	29.40	33.60	11.11	11.17	19.41	17.17
Palestinian	29.10	37.47	19.13	36.98	27.36	19.85	30.68	26.63	30.39	29.60	41.61	34.04	32.03	27.68	19.09	13.40	39.50	35.99	33.03	36.88	16.66	13.96	26.72	19.49
Qatar	30.38	38.78	20.36	39.20	28.23	20.31	32.75	28.65	27.40	32.79	45.10	35.11	30.39	28.84	19.73	14.97	40.91	37.28	31.73	41.08	20.46	16.40	21.39	21.27
Romania	37.99	45.45	29.17	47.27	36.15	26.89	38.63	38.40	36.87	37.51	51.59	41.46	37.42	38.05	29.59	21.31	45.29	46.56	39.40	45.15	28.11	24.61	33.35	27.34
Russian Fed.	55.65	62.93	46.99	66.83	54.42	40.47	58.63	56.03	54.28	52.19	70.53	57.86	52.17	58.71	50.17	34.27	64.57	65.76	54.92	59.40	48.57	40.03	53.39	42.18
Saudi Arabia	26.76	35.83	15.94	35.04	24.68	17.40	28.86	23.79	24.36	31.01	40.78	33.17	29.43	26.90	14.19	11.03	38.76	32.72	30.75	38.65	15.13	11.94	15.48	20.39
Singapore	72.68	77.90	66.46	81.48	72.46	59.52	77.39	70.95	71.61	69.91	82.53	75.98	66.49	77.50	68.11	55.83	81.40	75.55	72.03	77.55	68.70	61.73	71.02	59.29
Slovenia	46.64	54.67	37.07	57.01	45.08	33.25	51.83	38.17	46.36	53.37	64.57	48.94	45.07	46.64	40.30	26.99	56.18	46.90	48.42	63.41	42.53	24.04	43.49	39.42
Sweden	41.25	47.61	33.52	49.84	41.33	28.24	49.09	31.24	35.61	51.60	54.53	45.19	37.36	45.46	36.56	23.41	50.87	37.86	40.82	59.21	43.26	21.05	28.37	41.04
Syrian Arab R.	25.06	33.41	15.09	31.19	24.32	16.82	25.22	24.17	25.97	25.34	36.35	31.88	28.09	21.56	14.97	10.85	34.17	34.65	29.88	32.19	13.49	10.82	20.54	15.82
Thailand	31.14	39.38	21.32	38.34	30.24	21.53	33.37	26.82	29.57	36.57	43.78	36.73	32.29	25.38	22.21	15.83	41.74	36.37	33.44	45.20	21.27	16.02	24.20	24.58
Tunisia	29.31	36.42	20.84	37.20	28.24	18.93	32.65	24.91	29.70	31.35	40.54	33.66	28.21	26.31	21.54	14.02	38.22	35.30	32.35	36.72	21.42	13.27	26.02	23.89
Turkey	**37.89**	**45.62**	**28.68**	**44.02**	**37.55**	**29.01**	**36.03**	**34.81**	**37.28**	**46.13**	**49.11**	**43.92**	**39.77**	**34.03**	**29.68**	**23.31**	**44.27**	**43.82**	**41.05**	**54.52**	**27.53**	**24.72**	**32.06**	**34.48**
Ukraine	41.62	49.86	31.80	51.44	40.51	28.34	42.68	39.53	41.05	44.09	56.99	45.89	40.17	41.18	33.86	22.07	49.78	51.08	43.57	52.52	32.09	26.58	37.54	32.38
United Arab E.	36.78	44.89	27.12	48.30	33.64	24.20	40.94	34.06	32.55	39.55	52.28	39.83	33.78	37.89	25.97	19.14	48.24	44.22	36.54	46.78	28.03	22.68	27.01	29.50
United States	48.32	56.24	38.89	60.79	45.84	33.21	53.17	43.11	41.85	56.41	64.70	51.33	42.93	51.72	39.05	28.06	56.96	54.96	45.41	64.42	44.33	29.82	36.91	45.29
Avg.	**40.65**	**48.22**	**31.64**	**49.08**	**39.62**	**29.37**	**43.30**	**36.98**	**39.71**	**43.81**	**53.95**	**44.97**	**39.83**	**39.53**	**33.01**	**23.83**	**51.25**	**46.94**	**42.75**	**51.82**	**35.10**	**25.82**	**35.50**	**32.69**

Table A6. Rankings Based on Average Percent Correct for Each Item Type

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Armenia	21	20	21	19	22	23	20	10	23	36	23	23	23	20	30	23	17	11	27	27	20	14	19	33
Australia	12	12	12	13	11	10	12	18	13	8	12	12	10	11	11	11	14	19	10	10	12	12	15	9
Bahrain	32	34	32	31	34	31	34	31	36	27	33	33	34	29	34	28	34	33	35	35	33	31	34	24
Chile	36	35	36	37	30	34	35	39	33	26	34	31	30	37	31	36	36	40	29	29	31	36	36	28
Chinese Taipei	3	3	3	3	3	2	3	1	1	4	3	3	1	3	2	3	3	1	1	1	3	1	3	5
England	10	10	10	12	9	8	11	15	11	6	14	10	11	17	9	8	11	15	11	11	10	11	14	6
Finland	8	8	8	10	8	9	7	14	12	7	10	8	8	10	7	10	8	13	12	12	6	16	11	7
Georgia	27	27	28	27	26	29	27	24	30	28	26	26	29	28	26	29	28	22	28	28	26	26	33	35
Ghana	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Hong Kong	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	4	4	4	4	4	4	4	4
Hungary	11	11	9	9	10	11	10	11	8	12	9	14	13	9	10	9	13	16	8	8	9	10	10	10
Indonesia	40	39	40	40	39	40	39	40	40	38	38	40	37	40	38	40	39	39	38	38	40	40	40	38
Iran	31	31	33	35	31	27	33	30	26	31	30	30	26	34	32	26	32	30	26	26	34	32	24	37
Israel	7	7	7	8	7	7	8	7	14	11	11	7	9	7	8	7	7	7	14	14	8	7	12	11
Italy	15	13	17	15	14	14	15	16	7	18	15	9	7	15	16	16	10	14	7	7	18	17	8	18
Japan	5	5	5	5	5	5	5	5	5	3	5	5	5	5	5	4	5	5	5	5	5	5	5	3
Jordan	35	32	35	33	36	33	37	27	34	34	32	35	35	36	36	31	37	26	36	36	36	30	31	34
Kazakhstan	17	18	15	16	18	17	19	8	15	22	19	19	19	13	17	15	20	9	18	18	19	9	7	21
Korea	1	1	1	2	1	1	2	3	3	1	1	1	3	2	1	1	2	2	3	3	1	3	1	2
Lebanon	25	26	25	23	27	30	25	20	25	35	24	28	28	24	24	33	24	20	30	30	27	25	22	31
Lithuania	14	15	14	14	12	16	16	12	10	15	13	13	15	12	12	17	15	12	9	9	16	15	13	16
Macedonia	28	28	27	28	28	25	29	25	27	30	28	27	27	27	27	25	30	24	25	25	28	23	25	29
Malaysia	26	25	26	26	25	26	23	29	24	25	27	25	22	26	25	30	21	29	23	23	25	33	27	25
Morocco	41	41	41	41	41	41	41	41	39	41	41	41	41	41	37	41	41	41	41	41	39	41	39	41
New Zealand	16	16	16	17	16	15	17	21	16	10	16	16	16	18	15	13	18	25	15	15	15	18	17	14
Norway	20	23	20	24	20	21	18	34	21	14	21	21	24	19	19	20	23	36	21	21	17	27	23	12
Oman	39	40	39	38	40	39	40	38	38	39	39	39	40	39	41	39	40	38	40	40	41	38	38	39
Palestinian	34	33	34	34	35	35	32	33	29	37	35	34	32	31	35	35	31	32	32	32	35	34	29	36
Qatar	30	30	31	29	33	32	30	28	35	29	29	32	33	30	33	32	29	28	34	34	32	28	35	30
Romania	22	22	22	22	23	22	24	17	20	23	22	22	20	22	22	22	25	18	22	22	22	20	20	23
Russian Fed.	6	6	6	6	6	6	6	6	6	16	6	6	6	6	6	6	6	6	6	6	7	6	6	13
Saudi Arabia	37	37	37	36	37	37	36	37	41	33	36	37	36	32	40	37	33	37	37	37	37	37	41	32
Singapore	2	2	2	1	2	3	1	2	2	2	2	2	2	1	3	2	1	3	2	2	2	2	2	1
Slovenia	13	14	13	11	15	12	13	19	9	13	8	15	12	14	13	14	12	17	13	13	14	21	9	17
Sweden	19	19	18	20	17	20	14	26	22	17	18	18	21	16	18	18	16	27	20	20	13	24	26	15
Syrian Arab R.	38	38	38	39	38	38	38	36	37	40	40	38	39	38	39	38	38	35	39	39	38	39	37	40
Thailand	29	29	29	30	29	28	28	32	32	24	31	29	31	35	28	27	27	31	31	31	30	29	32	26
Tunisia	33	36	30	32	32	36	31	35	31	32	37	36	38	33	29	34	35	34	33	33	29	35	30	27
Turkey	**23**	**21**	**23**	**25**	**21**	**18**	**26**	**22**	**19**	**19**	**25**	**20**	**18**	**25**	**21**	**19**	**26**	**23**	**19**	**19**	**24**	**19**	**21**	**19**
Ukraine	18	17	19	18	19	19	21	13	18	20	17	17	17	21	20	21	19	10	17	17	21	13	16	20
United Arab E.	24	24	24	21	24	24	22	23	28	21	20	24	25	23	23	24	22	21	24	24	23	22	28	22
United States	9	9	11	7	13	13	9	9	17	9	7	11	14	8	14	12	9	8	16	16	11	8	18	8

Table A7. Dunnett's Test Results for Turkey and All Countries

Country	All	MC	CR	K.	A.	R.	N.	Alg.	Geo.	DC	MC-K	MC-A	MC-R	CR-K	CR-A	CR-R	MC-N	MC-Alg.	MC-Geo	MC-DC	CR-N	CR-Alg	CR-Geo	CR-DC
Korea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Singapore	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Chinese Tai.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016
Hong Kong	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011
Japan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
Russian Fed.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.619**	0.000	0.000	0.028	0.000	0.000	0.003	0.000	0.000	0.001	**0.808**	0.000	0.000	0.000	**0.897**
Israel	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.002	**0.053**	**0.284**	0.000	0.000	**0.590**	0.005	0.001	0.006	0.001	0.001	**0.366**	**0.253**	0.000	0.026	**0.241**	**0.837**
Finland	0.000	0.000	0.000	0.000	0.000	**0.191**	0.000	**1.000**	0.028	0.006	0.000	0.003	**0.560**	0.025	0.000	**0.328**	0.004	**0.776**	**0.203**	0.001	0.000	**1.000**	**0.234**	**0.394**
United States	0.000	0.000	0.001	0.000	0.003	**0.714**	0.000	0.043	**0.622**	**0.062**	0.000	0.038	**1.000**	0.006	**0.072**	**0.655**	0.008	0.004	**0.894**	**0.050**	0.000	**0.815**	**0.974**	**0.490**
England	0.000	0.000	0.000	0.000	0.000	**0.139**	0.000	**0.804**	0.019	0.005	0.000	0.029	**0.800**	**0.285**	0.001	**0.140**	0.015	**0.877**	**0.121**	0.002	0.000	**0.972**	**0.263**	**0.263**
Hungary	0.000	0.000	0.000	0.000	0.001	**0.289**	0.000	**0.988**	0.005	**0.291**	0.000	**0.245**	**0.985**	0.023	0.002	**0.232**	0.032	**0.977**	**0.061**	**0.359**	0.000	**0.977**	**0.134**	**0.732**
Australia	0.000	0.000	0.000	0.000	0.001	**0.250**	0.000	**1.000**	0.029	0.024	0.000	**0.074**	**0.696**	0.035	0.013	**0.368**	0.040	**1.000**	**0.113**	0.004	0.000	**1.000**	**0.411**	**0.576**
Slovenia	0.000	0.000	0.011	0.000	0.010	**0.700**	0.000	**1.000**	0.015	**0.379**	0.000	**0.351**	**0.881**	**0.208**	0.025	**0.923**	0.016	**0.999**	**0.252**	**0.106**	0.001	**1.000**	**0.095**	**1.000**
Lithuania	0.000	0.000	0.011	0.000	0.001	**1.000**	0.012	**0.999**	0.017	**0.469**	0.000	**0.074**	**1.000**	**0.056**	0.016	**1.000**	**0.372**	**0.465**	**0.112**	**0.348**	0.006	**1.000**	**0.262**	**0.980**
Italy	0.000	0.000	**0.113**	0.000	0.005	**0.815**	0.011	**1.000**	0.001	**0.980**	0.000	0.024	**0.498**	**0.317**	**0.152**	**1.000**	0.008	**0.888**	0.014	**0.808**	**0.188**	**1.000**	**0.095**	**1.000**
New Zealand	0.013	**0.128**	0.039	0.002	**0.110**	**0.954**	0.027	**1.000**	**0.551**	**0.227**	0.003	**0.899**	**1.000**	**0.405**	**0.109**	**0.810**	**0.711**	**1.000**	**0.878**	**0.108**	0.006	**1.000**	**0.935**	**0.928**
Kazakhstan	**0.397**	**1.000**	**0.106**	**0.052**	**0.631**	**1.000**	**0.547**	**0.501**	**0.263**	**0.453**	**0.473**	**1.000**	**1.000**	**0.193**	**0.255**	**0.999**	**0.931**	**0.229**	**1.000**	**0.063**	**0.493**	**0.865**	0.041	**1.000**
Ukraine	**0.381**	**0.280**	**0.972**	0.036	**0.873**	**1.000**	**0.827**	**0.643**	**0.855**	**1.000**	0.031	**1.000**	**1.000**	**0.892**	**0.935**	**1.000**	**0.809**	**0.182**	**1.000**	**1.000**	**0.968**	**1.000**	**0.914**	**1.000**
Sweden	**0.728**	**1.000**	**0.452**	**0.067**	**0.577**	**1.000**	0.005	**0.533**	**1.000**	**0.751**	**0.425**	**1.000**	**1.000**	**0.351**	**0.353**	**1.000**	**0.557**	**0.516**	**1.000**	**0.849**	0.001	**0.996**	**1.000**	**0.978**
Norway	**1.000**	**1.000**	**0.959**	**0.430**	**1.000**	**1.000**	**0.160**	0.010	**1.000**	**0.388**	**0.987**	**1.000**	**0.993**	**0.601**	**0.934**	**1.000**	**1.000**	0.014	**1.000**	**0.417**	0.010	**0.296**	**1.000**	**0.861**
Armenia	**1.000**	**0.961**	**1.000**	**0.835**	**1.000**	**0.911**	**0.617**	**0.969**	**1.000**	0.000	**1.000**	**0.976**	**0.997**	**0.754**	**1.000**	**0.975**	**0.584**	**0.314**	**0.565**	0.000	**0.901**	**1.000**	**1.000**	**0.157**
Romania	**0.999**	**1.000**	**1.000**	**0.991**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.188**	**1.000**	**0.995**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.078**	**1.000**	**1.000**	**1.000**	**0.946**
Turkey	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.902**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
U. Arab E.	**0.866**	**0.997**	**0.985**	**0.917**	**0.528**	**0.530**	**1.000**	**1.000**	**0.574**	**0.506**	**0.971**	**0.626**	**0.750**	**1.000**	**0.982**	**0.810**	**0.993**	**1.000**	**0.866**	**0.224**	**1.000**	**1.000**	**0.960**	**1.000**
Lebanon	0.008	**0.073**	**0.140**	**1.000**	**0.100**	0.049	**1.000**	**0.995**	**0.709**	0.000	**1.000**	**0.054**	**0.578**	**1.000**	**0.935**	**0.053**	**1.000**	**1.000**	**0.308**	0.000	**0.982**	**0.671**	**1.000**	**0.213**
Malaysia	0.022	**0.376**	**0.078**	**1.000**	**0.306**	**0.161**	**1.000**	0.025	**0.988**	**0.092**	**1.000**	**0.526**	**0.999**	**1.000**	**0.785**	**0.070**	**0.993**	**0.342**	**1.000**	0.049	**1.000**	**0.062**	**0.994**	**0.673**
Georgia	0.003	**0.125**	0.024	**1.000**	**0.102**	**0.059**	**0.998**	**0.935**	**0.133**	0.007	**1.000**	**0.274**	**0.569**	**1.000**	**0.466**	**0.071**	**1.000**	**1.000**	**0.554**	**0.013**	**1.000**	**0.564**	**0.398**	**0.128**
Macedonia	0.006	0.040	**0.114**	**0.998**	0.042	**0.275**	**0.581**	**1.000**	**0.661**	0.003	**0.992**	**0.136**	**0.680**	**1.000**	**0.340**	**0.425**	**0.996**	**1.000**	**0.753**	0.001	**0.834**	**0.998**	**1.000**	**0.283**
Thailand	0.000	0.002	0.005	**0.169**	0.014	**0.072**	**0.647**	**0.053**	**0.065**	**0.102**	**0.374**	0.048	**0.455**	**0.658**	**0.257**	**0.120**	**1.000**	**0.160**	**0.220**	**0.078**	**0.676**	**0.146**	**0.493**	**0.610**
Qatar	0.000	0.000	0.001	**0.816**	0.000	0.020	**0.428**	**0.071**	0.006	0.004	**0.854**	0.006	**0.189**	**0.999**	0.045	**0.057**	**1.000**	**0.381**	**0.067**	0.002	**0.501**	**0.218**	**0.145**	**0.241**
Iran	0.000	0.000	0.001	**0.282**	0.001	**0.115**	**0.098**	0.022	**0.677**	0.002	**0.501**	0.013	**0.718**	**0.759**	**0.066**	**0.127**	**0.909**	**0.272**	**0.676**	0.001	**0.140**	**0.071**	**1.000**	**0.124**
Bahrain	0.000	0.000	0.001	**0.504**	0.000	0.028	**0.084**	0.013	0.004	0.046	**0.304**	0.002	**0.161**	**1.000**	0.028	**0.096**	**0.692**	**0.112**	**0.050**	0.003	**0.199**	**0.083**	**0.148**	**0.896**
Tunisia	0.000	0.000	0.001	0.023	0.000	0.003	**0.266**	0.007	**0.073**	0.001	0.013	0.001	**0.050**	**0.813**	**0.171**	0.022	**0.686**	**0.065**	**0.106**	0.000	**0.709**	0.016	**0.830**	**0.518**
Palestinian	0.000	0.000	0.000	**0.140**	0.000	0.011	**0.058**	0.010	**0.135**	0.000	**0.078**	0.001	**0.403**	**0.984**	0.026	0.011	**0.939**	**0.170**	**0.169**	0.000	**0.058**	0.041	**0.932**	**0.126**
Jordan	0.000	0.000	0.000	**0.217**	0.000	0.018	0.013	**0.087**	0.019	0.000	**0.372**	0.001	**0.141**	**0.746**	0.016	**0.064**	**0.410**	**0.690**	0.048	0.000	0.042	**0.137**	**0.519**	**0.145**
Chile	0.000	0.000	0.000	**0.081**	0.002	0.012	**0.168**	0.000	0.039	**0.087**	**0.104**	0.011	**0.537**	**0.732**	**0.110**	0.008	**0.534**	0.001	**0.473**	**0.144**	**0.571**	0.010	**0.136**	**0.389**
Saudi Arabia	0.000	0.000	0.000	**0.064**	0.000	0.000	0.018	0.000	0.000	0.001	0.032	0.000	**0.110**	**0.940**	0.000	0.001	**0.810**	0.009	0.029	0.000	0.017	0.007	0.002	**0.178**
Syrian A. R.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.046	**0.222**	0.000	0.000	**0.073**	**0.059**	0.013	0.000	0.004	0.002	**0.091**	0.024
Oman	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.037	**0.178**	0.000	0.000	**0.062**	0.003	0.008	0.000	0.000	0.003	0.045	0.047
Indonesia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	**0.090**	**0.063**	0.000	0.000	**0.066**	0.001	0.018	0.000	0.001	0.001	0.008	**0.094**
Morocco	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.033	0.000	0.000	0.006	0.000	0.007	0.000	0.002	0.000	0.039	0.006
Ghana	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Inter. avg.	**0.743**	**0.919**	**0.971**	**0.616**	**1.000**	**1.000**	**0.276**	**1.000**	**1.000**	**1.000**	**0.680**	**1.000**	**1.000**	**0.999**	**1.000**	**1.000**	**0.594**	**1.000**	**1.000**	**1.000**	**0.558**	**1.000**	**1.000**	**1.000**

Bold ones are not significantly different with Turkey ($p < 0.05$)

REFERENCES

Acar, T. (2012). The position of Turkey among OECD member and candidate countries according to PISA 2009 results. *Educational Sciences: Theory and Practice, 12*(4), 2567-2572.

Adams, R., Berezner, A., & Jakubowski, M. (2010). Analysis of PISA 2006 preferred items ranking using the percent-correct method. *OECD Education Working Papers*, No. 46, OECD Publishing.

Akpınar, E. (2003). Ortaöğretim coğrafya dersleri yazılı sınav sorularının bilişsel düzeyleri. *Erzincan Eğitim Fakültesi Dergisi*, *5*(1), 13-21.

Arı, A., & Gökler, Z. S. (2012). *İlköğretim fen ve teknoloji dersi kazanımları ve SBS sorularının yeni Bloom taksonomisine göre değerlendirilmesi*. X. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Niğde.

Aydoğdu İskenderoğlu, T., & Baki, A. (2011). İlköğretim 8. sınıf matematik ders kitabındaki soruların PISA matematik yeterlik düzeylerine göre sınıflandırılması. *Eğitim ve Bilim*, *36*(161), 287-301.

Ayvacı H. S., & Şahin C. (2009). Fen bilgisi öğretmenlerinin ders sürecinde ve yazılı sınavlarda sordukları soruların bilişsel seviyelerinin karşılaştırılması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *22*, 241-255.

Azar, A. (2005). Analysis of Turkish high-school physics examination questions and university entrance exams questions according to Blooms' taxonomy. *Türk Fen Eğitimi Dergisi*, *2*(2), 144–150.

Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education*, *25*(1), 31-36.

Baysen, E. (2006). Öğretmenlerin sınıfta sordukları sorular ile öğrencilerin bu sorulara verdikleri cevapların düzeyleri. *Kastamonu Eğitim Dergisi*, *14*(1), 21-28.

Beaton, A. E. (1998). Comparing cross-national student performance on TIMSS using different test items. *International Journal of Educational Research*, *29*(6), 529-542.

Beckmann, S. (2004). Solving algebra and other story problems with simple diagrams: A method demonstrated in grade 4-6 texts used in Singapore. *Mathematics Educator*, *14*(1), 42-46.

Bekdemir, M., & Selim, Y. (2008). Revize edilmiş Bloom Taksonomisi ve cebir öğrenme alanı örneğinde uygulaması. *Erzincan Eğitim Fakültesi Dergisi*, *10*(2), 185-196.

Bennett, R. E., Ward, W. C., Rock, D. A., & LaHart, C. (1990). *Toward a framework for constructed-response items*. Princeton, NJ: Educational Testing Service.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals, handbook I: The cognitive domain*. New York: David Mckay Company Inc.

Bush, M. (1999). *Alternative marking schemes for on-line multiple-choice tests.* Paper presented at the Seventh Annual Conference on the Teaching of Computing, Belfast, Ireland. Retrieved from http://www.caacentre.ac.uk/dldocs/BUSHMARK.PDF

Büyüköztürk, Ş., Çakan, M., Tan, Ş., & Atar, H. Y. (2014). *TIMSS 2011 ulusal matematik ve fen raporu: 8. sınıflar*. Ankara: İşkur Matbaacilik Ltd. Şti.

Ceylan, E. (2013). Investigating science content and cognitive domain scores with regard to low-and high-performing schools in Turkey. *Journal of Education and Future*, *4*, 35-50.

Coşar, N. (2010*). İlköğretim 6. sınıf matematik ders kitaplarındaki problemlerin analizi* (Master's thesis). Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa, Turkey.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *37*(2), 99-114.

Çepni, S., Ayvacı, H., & Keleş, E. (2001). Okullarda ve lise giriş sınavlarında sorulan fen bilgisi sorularının Bloom taksonomisine göre karşılaştırılması. *Yeni Bin Yılın Başında Fen Bilimleri Eğitimi Sempozyumu* (pp. 144-150). İstanbul: Maltepe Üniversitesi Eğitim Fakültesi.

Çevik, Ş. (2010). *Ortaöğretim 9, 10 ve 11. sınıf fizik ders kitaplarında bulunan sorular ile 2000-2008 yılları arasında öğrenci seçme ve yerleştirme sınavlarında sorulan fizik sorularının Bloom taksonomisi açısından*

incelenmesi ve karşılaştırılması (Unpublished master's thesis). Dicle Üniversitesi Fen Bilimleri Enstitüsü, Diyarbakır, Turkey.

Darling-Hammond, L. (1994). Performance assessment and educational equity. *Harvard Educational Review, 64*(1), 5-30.

Delil, A., & Tetik, B. Y. (2015). 8. sınıf merkezi sınavlardaki matematik sorularının TIMSS-2015 bilişsel alanlarına göre analizi. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 13*(4), 165-184.

Delil, H. (2006). *An analysis of geometry problems in 6-8 grades Turkish mathematics textbooks* (Master's thesis). Middle East Technical University, Ankara, Turkey.

Dindar, H., & Demir, M. (2006). Beşinci sınıf öğretmenlerinin fen bilgisi dersi sınav sorularının Bloom taksonomisine göre değerlendirilmesi. *Gazi Eğitim Fakültesi Dergisi, 26*(3), 87-96.

Dogan, E., & Tatsuoka, K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics, 68*(3), 263-272.

Dursun, A., & Aydin-Parim, G. (2014). YGS 2013 matematik soruları ile ortaöğretim 9. sınıf matematik sınav sorularının Bloom taksonomisine ve öğretim programına göre karşılaştırılması [A comparison of mathematics questions in the 2013-YGS examination and the teacher made ninth-grade class, using Bloom's taxonomy]. *Eğitim Bilimleri Araştırma Dergisi - Journal of Educational Sciences Research, 4*(1), 17-37.

Elley, W. B., & Mangubhai, F. (1992). Multiple-choice and open-ended items in reading tests: Same or different? *Studies in Educational Evaluation, 18*, 191-199.

Erman, E. (2008). *2003-2006 yılları arasında yapılan ortaöğretim kurumlarına öğrenci seçme sınavında yer alan tarih bilimi sorularının Bloom taksonomisine göre değerlendirilmesi* (Yüksek Lisans Tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara, Turkey.

Eroğlu, D., & Sarar Kuzu, T. (2014). Türkçe ders kitaplarındaki dilbilgisi kazanımlarının ve sorularının yenilenmiş Bloom taksonomisine göre değerlendirilmesi. *Başkent University Journal of Education, 1*(1), 72-80.

Farthing, D. W., Jones, D. M., & McPhee, D. (1998). *Permutational multiple-choice questions: An objective and efficient alternative to essay-type examination questions*. Paper presented at the ITiCSE '98, Dublin, Ireland. Retrieved from http://bioinfo.uib.es/~joemiro/semdoc/p81-farthing.pdf

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (6th ed.). London: SAGE Publications Ltd.

Foster, D. (2005). *Multiple-choice questions are the answer.* Retrieved from www.certmag.com/read.php?in=1171

Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013a). *TIMSS 2011 user guide for the international database*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/timss2011/international-database.html

Foy, P., Brossman, B., & Galia, J. (2013b). *Methods and procedures in TIMSS and PIRLS 2011: Scaling the TIMSS and PIRLS 2011 achievement data*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/methods/t-achievement-scales.html

Foy, P., Martin, M. O., Mullis, I. V. S., & Stanco, G. (2013c). *Methods and procedures in TIMSS and PIRLS 2011: Reviewing the TIMSS and PIRLS 2011 achievement item statistics*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/methods/t-achievement-scales.html

Gezer, M., Şahin, İ. F., Sünkür, M. Ö., & Meral, A. (2014). 8. sınıf Türkiye Cumhuriyeti inkılap tarihi ve Atatürkçülük dersi öğretim programı kazanımlarının revize edilmiş Bloom taksonomisine göre değerlendirilmesi. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, *3*(1), 433-455.

Gökulu, A. (2015). Fen ve teknoloji öğretmenlerinin yazılı sınav soruları ile TEOG sınavlarında sorulan fen ve teknoloji sorularının yenilenmiş Bloom taksonomisine göre incelenmesi. *Route Educational and Social Science Journal*, *2*(2), 434-446.

Gravetter, F. J., & Wallnau, L. B. (2014*). Essentials of statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Güler, G., Özdemir, E., & Dikici, R. (2012). İlköğretim matematik öğretmenlerinin sınav soruları ile SBS matematik sorularının Bloom taksonomisine göre karşılaştırmalı analizi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *14*(1), 41-60.

Gündüz, Y. (2009). İlköğretim 6, 7 ve 8. sınıf fen ve teknoloji sorularının ölçme araçlarına ve Bloom'un bilişsel alan taksonomisine göre analizi. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *6*(2), 150-165.

Güner, N. (2015). 6.-8. Sınıf matematik ders kitaplarındaki geometri, veri ve olasılık sorularının TIMSS bilişsel düzeylerine göre sınıflandırılması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *37*(1), 77-90.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education*, *62*(2), 143-157.

Hastedt, D. & Sibberns, H. (2005). Differences between multiple choice items and constructed response items in the IEA TIMSS surveys. *Studies in Educational Evaluation*, *31*, 145-161.

Hastedt, D. (2004). Differences between multiple-choice and constructed response items in *PIRLS 2001*. In Proceedings of the 1st IEA International Research Conference. Nicosia, Cyprus.

Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Boston, MA: Pearson.

Incikabi, L. (2012). After the reform in Turkey: A content analysis of SBS and TIMSS assessment in terms of mathematics content, cognitive domains, and item types. *Education as Change*, *16*(2), 301-312.

International Association for the Evaluation of Educational Achievement (2013a). TIMSS 2011 Released Items. Retrieved from http://timss.bc.edu/timss2011/international-released-items.html

International Association for the Evaluation of Educational Achievement (2013b). TIMSS 2011 International Database. Retrieved from http://timssandpirls.bc.edu/timss2011/international-database.html

Jennings, S., & Bush, M. (2006). A comparison of conventional and liberal (free-choice) multiple choice tests. *Practical Assessment Research & Evaluation*, *11*(8).

Joncas, M. (2013). *Methods and procedures in TIMSS and PIRLS 2011:TIMSS 2011 target population*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/methods/t-sample-design.html

Joncas, M., & Foy, P. (2013). *Methods and procedures in TIMSS and PIRLS 2011: sample design in TIMSS and PIRLS*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/methods/t-sample-design.html

Kablan, Z., Baran, T., & Hazer, Ö. (2013). İlköğretim matematik 6-8 öğretim programında hedeflenen davranışların bilişsel süreçler açısından incelenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, *14*(1), 347-366.

Kaleli-Yılmaz, G., & Hanci, A. (2016). Examination of the 8th grade students' TIMSS mathematics success in terms of different variables. *International Journal of Mathematical Education in Science and Technology*, *47*(5), 674-695.

Karakuş, F. (2010). Ortaöğretim matematik dersi öğretim programında yer alan alternatif ölçme ve değerlendirme yaklaşımlarına yönelik öğretmen görüşleri. *Türk Eğitim Bilimleri Dergisi*, *8*(2), 457-488.

Karaman, İ. (2005). Erzurum ilinde bulunan liselerdeki fizik sınav sorularının Bloom taksonomisinin basamaklarına göre analizi. *Gazi Eğitim Fakültesi Dergisi*, *25*(1), 77-90.

Karamustafaoğlu, S., Sevim, S., Karamustafaoğlu, O., & Çepni, S. (2003). Analysis of Turkish high school chemistry-examination questions according to Bloom's taxonomy. *Chemistry Education*: *Research and Practice*, *4*(1), 25-30.

Kastner, M., & Stangl, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter? *Procedia Social and Behavioral Sciences*, 12, 263-173.

Keleş, T., & Hacısalihoğlu-Karadeniz, M. (2015). An analysis of mathematics and geometry questions in OSS, YGS and LYS according to the revised Bloom's taxonomy between 2006-2012 years. *Turkish Journal of Computer and Mathematics Education*, *6*(3), 532-552.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, N.J: Pearson Prentice Hall.

Keskin, M. Ö., & Aydın, S. (2011). Seviye belirleme sınavı 6. sınıf fen ve teknoloji testinde çıkan biyoloji sorularının revize edilmiş taksonomiye göre incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 31*(3), 727-742.

Koç, H., Sönmez, Ö. F., & Çiftçi, T. (2013). ÖSS, YGS ve LYS sınavlarındaki coğrafya sorularının Bloom taksonomisi bilişsel alan düzeyi açısından analizi. *Karadeniz Araştırmaları, 36,* 257-275.

Köğce, D. (2005). *ÖSS sınavı matematik soruları ile liselerde sorulan yazılı sınav sorularının Bloom taksonomisine göre karşılaştırılması* (Unpublished PhD thesis). Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü. Trabzon, Turkey.

Köğce, D., & Baki, A. (2009). Matematik öğretmenlerinin yazılı sınav soruları ile ÖSS sınavlarında sorulan matematik sorularının Bloom taksonomisine göre karşılaştırılması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *26*(26), 70-80.

Kuran, K., & Kanatlı, F. (2009). Alternatif ölçme değerlendirme teknikleri konusunda sınıf öğretmenlerinin görüşlerinin değerlendirilmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *6*(12), 209-234.

Küçük, E. E. (2008). İlköğretim 6. sınıf Türkçe çalışma kitabındaki soruların eleştirel düşünme açısından incelenmesi. *e-Journal of New World Sciences Academy Social Sciences, 3,* (3), 492-504.

Lapointe A. E., Askew, J. M., & Mead, N. A. (1992). *Learning science* (pp.11). Princeton, NJ: Educational Testing Service.

Lissitz, R, & Hou, X. (2008). *Multiple choice items and constructed response items: does it matter?* Retrieved from http://www.education.umd.edu/EDMS/MARCES/multiple choice items and constructed response items.pdf.

Livingston, S. A. (2009) Constructed-response test questions: Why we use them; how we score them. *ETS: R&D Connections*, *11*, 1-8.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*(3), 234-250.

Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011: TIMSS and PIRLS stratified two-stage cluster sample design*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved March 15, 2016, from http://timssandpirls.bc.edu/methods/t-sample-design.html

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist, 34*(4), 207-218.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's repeat of the third international mathematics and science study at the eighth grade*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College Chestnut Hill, MA, USA. Retrieved March 15, 2016, from https://timssandpirls.bc.edu/timss1999i/math_achievement_report.html

Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College Chestnut Hill, MA, USA. Retrieved March 15, 2016, from https://timssandpirls.bc.edu/TIMSS2007/mathreport.html

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College Chestnut Hill, MA, USA. Retrieved March 15, 2016, from http://timssandpirls.bc.edu/timss2011/international-results-mathematics.html

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks.* TIMSS & PIRLS International Study Center Lynch School of Education, Boston College, Chestnut Hill, MA, USA. Retrieved March 15, 2016, from http://timssandpirls.bc.edu/timss2011/frameworks.html

Nixon, R. S., & Barth. K. N. (2014). A comparison of TIMSS items using cognitive domains. *School Science and Mathematics, 114*(2), 65-75.

O'Leary, M. (2001). Item format as a factor affecting the relative standing of countries in the third international mathematics and science study (TIMSS). *Educational Studies*, *20*(1), 153-177.

O'Leary, M. (2002). Stability of country rankings across item formats in the third international mathematics and science study. *Educational Measurement: Issues and Practice, 21*(4), 27-38.

Özcan, S., & Oluk, S. (2007). İlköğretim fen bilgisi derslerinde kullanılan soruların Piaget ve Bloom taksonomisine göre analizi. *Ziya Gökalp Eğitim Fakültesi Dergisi, 8,* 61-68.

Özmen, H., & Karamustafaoğlu, O. (2006). Lise 2. sınıf fizik-kimya sınav sorularının ve öğrencilerin enerji konusundaki başarılarının bilişsel gelişim seviyelerine göre analizi. *Kastamonu Eğitim Dergisi, 14*(1), 91-100.

Öztürk, N., Yalvaç Hastürk, N.G., & Demir, R. (2013). İlköğretim 4-5. sınıf fen ve teknoloji dersi öğretim programlarındaki ölçme ve değerlendirme yöntemlerine ilişkin öğretmen görüşleri. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, *20,* 25-36.

Rovai, A. P., Baker, J. D., & Ponton, M. K. (2013). *Social sciences research design and statistics: A practitioner's guide to research methods and IBM SPSS analysis* (2nd ed.). Chesapeake, VA: Watertree Press.

Sesli, A. (2007). *Biyoloji öğretmenlerinin yazılı sınav soruları ile ÖSS sorularının Bloom taksonomisine göre karşılaştırmalı analizi* (Yayınlanmamış yüksek lisans tezi). Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Trabzon, Turkey.

Seven, S. (2001). *İlköğretim sosyal bilgiler ders kitapları hakkında öğretmen ve öğrenci görüşleri*. (Unpublished master's thesis). Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa, Turkey.

Tanık, N., & Saraçoğlu, S. (2011). Fen ve teknoloji dersi yazılı sorularının yenilenmiş Bloom taksonomisine göre incelenmesi. *TÜBAV Bilim Dergisi, 4*(4), 235-246.

Tekin, H. (2003). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları.

Temizkan, M., & Sallabaş, M. E. (2011). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması. *Dumlupınar Sosyal Bilimler Dergisi*, *30*, 207-220.

Uğurel, I., Moralı, H. S., & Kesgin, Ş. (2012). OKS, SBS ve TIMSS matematik sorularının 'MATH Taksonomi' çerçevesinde karşılaştırmalı analizi. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi*, *11*(2), 423-444.

Uzun, S., Bütüner, S. Ö., & Yiğit, N. Y. (2010). A comparison of the results of TIMSS 1999-2007: The most successful five countries-Turkey sample. *İlköğretim Online*, *9*(3), 1174-1188.

Ünal, H., Demir, I., & Kılıç, S. (2011). Teachers' professional development and students' mathematics performance: Findings from TIMSS 2007. *Procedia Social and Behavioral Sciences*, *15*, 3252-3257.

Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education*, *54*, 455-461.

Walstad, W. B., & Becker, W. E. (1994). Achievement differences on multiple choice and essay tests in economics. *American Economic Review, 84*(2), 192-196.

Wolf, R. M. (1998). Validity issues in international assessments. *International Journal of Educational Research, 29,* 491-501.

Zhu, Y., & Fan, L. (2006). Focus on the representation of problem types in intended curriculum: A comparison of selected mathematics textbooks from Mainland China and the United States. *International Journal of Science and Mathematics Education, 4*(4), 609-626.