Kastamonu Education Journal, 2023, Vol. 31, No:1, 133-142
doi: 10.24106/kefdergi.1246453

| Research Article / Araştırma Makalesi |

A Mixture Rasch Model Analysis of Mathematics Achievement

Matematik Başarısının Karma Rasch Model ile Analizi

Halime YILDIRIM HOŞ[1], **Menekşe UYSAL SARAÇ**[2]

Keywords

1. Mixed item response theory
2. Mixture Rasch model
3. Mathematics achievement

Anahtar Kelimeler

1. Karma madde tepki kuramı
2. Karma Rasch model
3. Matematik başarısı

Received/Başvuru Tarihi
29.11.2021

Accepted / Kabul Tarihi
20.09.2022

Abstract

Purpose: This study aims to determine distinct latent classes in 8th-grade students' mathematics achievement

Design: The research study group consisted of 435 students who received the first booklet at the TIMSS 2015 8th grade mathematics achievement test. A mixture of Rasch model analysis was used to find the best-fitting model.

Findings: When model fit indices were evaluated, the model fitting the data was found to be MRM with the two latent classes. According to this model, students in latent class 2 are more successful in mathematics than students in latent class 1. The vast majority of the items are found easier in latent class 2.

Highlights: In addition, the cognitive domain of the more accessible items for students in both classes is “knowing, while items that are difficult for both groups are in the “applying” cognitive domain.

Öz

Çalışmanın amacı: Bu çalışmanın amacı, 8. sınıf öğrencilerinin matematik başarılarındaki farklı örtük sınıfları belirlemektir.

Materyal ve Yöntem: Araştırmanın çalışma grubu, TIMSS 2015 8. sınıf matematik başarı testinde ilk kitapçığı alan 435 öğrenciden oluşmaktadır. Veriye en uygun modeli bulmak için karma Rasch model (KRM) analizi kullanılmıştır.

Bulgular: . Model uyum indeksleri değerlendirildiğinde, verilere en iyi uyan modelin iki örtük sınıflı KRM olduğu görülmüştür. Bu modele göre, 2. örtük sınıftaki öğrenciler matematikte 1. örtük sınıftaki öğrencilere göre daha başarılıdır. Maddelerin büyük çoğunluğu 2. sınıftaki öğrencilere daha kolay gelmiştir.

Önemli Vurgular: Ayrıca, her iki sınıfta da öğrencilere kolay gelen maddelerin bilişsel alanı "bilmek" iken, iki grup için zor olan maddeler “uygulama” bilişsel alanındadır.

[1] **Corresponded Author,** İstanbul Medeniyet University, Faculty of Education, Educational Sciences, İstanbul, TURKEY; https://orcid.org/0000-0001-6479-4469
[2] Çankırı Karatekin University, Faculty of Letters, Educational Sciences, Çankırı, TURKEY; https://orcid.org/0000-0002-9847-1243

Citation/Alıntı: Yıldırım Hoş, H. & Uysal Saraç, M. (2023). A Mixture Rasch Model Analysis of Mathematics Achievement, *Kastamonu Education Journal, 31*(1), 143-152. doi: 10.24106/kefdergi.1246453

INTRODUCTION

Standard Rasch model (Rasch, 1960/1980) is unidimensional and as an assumption in this model people on the same ability level are comparable, that is, examinees with the same parameters are expected to have similar comments about their ability levels, skills and mental processes. On the other hand, studies show that persons at the same trait level, having the same measures on the construct, may use different processes, strategies, and operations to access the solutions (Baghaei & Carstensen, 2013; Mislevy & Verhelst, 1990; Cohen & Bolt, 2005; Ölmez & Cohen, 2018; Rost, 1990). This means that the trait and its meaning differentiate individuals depending on the processes and strategies they use to solve the items or problems; therefore, this can be seen as a significant threat to construct validity (Baghaei & Carstensen, 2013).

Relevant classes are determined according to the processes, strategies and operations individuals use to answer the items. If examinees with similar trait levels have different interpretations regarding the mechanisms and strategies, comparing people on a common ability continuum is not applicable. In other words, the underlying constructs measured by the instrument are different for different subpopulations of individuals, and the comparison of these individuals from different classes is not justifiable. (Embretson, 2007; Glück & Spiel, 2007; Rost, Carstensen, & von Davier, 1997).

In cognitive psychology, people with the same performance level may have qualitatively different strategies which underlie their performance. These qualitative differences among the individuals mean that construct representation also differs for them. When two or more latent class exists, different strategies, components, and knowledge structures in performance may lead to different orders of item difficulties. That is the nature of the trait bases on the class to which the person belongs. In this situation, the test measuring the construct may have different correlation coefficients from other measures. So, group membership is a moderator variable which is evidence for the changing meaning of the trait (Embretson, 2007).

Mixture item response theory models (MixIRT) combine item response theory (IRT) and latent class analysis (LCA), and they are used in analyzing item response data that may violate underlying assumptions of both or any of the modeling approaches (Rost, 1990). The most commonly used MixIRT model is the mixture Rasch model (MRM), and the underlying assumptions of Rasch and LCA models are briefly mentioned following (Li, Jiao, & Macready, 2016).

In the Rasch model, item difficulties must be constant for all individuals as a vital requirement, but in some cases, it may be violated for some items. Besides, items may have differing difficulties for different subgroups of individuals because the strategies of solving items they use are different, or they may have differentiating cognitive structures. Hence, a latent class approach and item analysis might be preferable. Also, in LCA models, the requirement is holding response probabilities for all individuals in the same latent class. However, various latent classes are necessary for determining individual ability differences for every cognitive construct or solution process. Accordingly, a generalized LCA model which allows varying ability levels within the latent class would be agreeable (Rost, 1990).

This study aims to determine the latent classes using the MRM and based on the TIMSS 2015 mathematical data. So, the items' properties according to these latent classes can be examined. This way, latent classes and students' latent class memberships can be determined simultaneously without solid assumptions about the Standard Rasch Model and Latent Class Analysis.

TIMSS (Trends in International Mathematics-Science Study) is a screening study conducted by the International Association for the Evaluation of Educational Achievement (IEA) in science and mathematics fields every four years. The TIMSS research aims to determine the knowledge and skills of the countries' 4th and 8th-grade students. Turkey participates in TIMSS to evaluate students' success in international fields and to compare its current educational system with those in other countries (Ministry of National Education [MoNE], 2016).

TIMSS provides participant countries various resources so the interferences in their studies and educational programs can be tested and commented on results. TIMSS includes achievement tests and questionnaires that students are asked about their school characteristics and teachers. These tests and questionnaires are used to collect information about student's performance in science and mathematics, educational systems, student characteristics, and characteristics of teachers and schools.

As a large-scale assessment study, TIMSS provides opportunities for researchers to determine the relationship between countries' education systems and students' success, compare countries in terms of their math and science success levels, and determine students' success and factors that affect student's success (MoNE, 2016; Tavşancıl & Yalcin, 2015). So, the TIMSS enables countries to evaluate students' mathematics achievement levels in terms of process and class level.

In the following, the Standard Rasch model and MRM used in the analysis are briefly explained.

The Rasch Model

For this purpose, item response models were developed, which differ according to several parameters used to describe items. In IRT models, the probability of an examinee answering an item correctly depending on the latent ability underlying the construct and characteristics of the item is determined (Hambleton, Swaminathan & Rogers, 1991). In the Rasch Model, one of the most commonly used IRT models, individual abilities and item difficulties are located on a standard scale. Therefore, an individual's probability of answering an item correctly is estimated by a function of the difference between the individual's latent ability and the difficulty of that item. (Rasch, 1960/1980). The mathematical representation of this function is given as follows:

$$P(x = 1|\theta j, \beta i) = \frac{\exp(\theta j - \beta i)}{1 + \exp(\theta j - \beta i)}$$

Where, examinee j's latent ability parameter is θ_j and the difficulty of item parameter is β_i. It is expected that the probability of giving correct answers is relatively higher for individuals with higher ability levels and/or easier items.

The Mixture Rasch Model

MRM combines the Rasch model and latent class analysis and it defines latent classes that the Rasch model is applied individually (Rost, 1990). In MRM it is assumed that the population is consist of different latent classes and for each classes item parameters may differ.

When items' difficulty patterns consistently differ among classes of population, MRM can fit the data. Rasch model is unidimensional and when it does not fit all the population, MRM allowing item parameters to differentiate among classes of this population can be chosen (Rost, 1990; Rost & von Davier, 1995). When the population is heterogeneous unavoidably, instead of rejecting the entire Rasch unscalable dataset, by considering different cognitive strategies for latent classes of population MRM may be used (Rost, 1990).

In MRM the correct response probability to an item is a function of both the examinee's ability which is a continuous variable and the categorical grouping variable regarding the set of strategies used for solution.

In MRM, for each latent class separate item difficulty parameters and for each examinee probability of assignment to particular latent class are estimated. Dichotomous i.e. scored in two categories MRM's mathematical representation is given as follows:

$$P(xij = 1|g, \theta jg,) = \frac{\exp(\theta jg - \beta ig)}{1 + \exp(\theta jg - \beta ig)}$$

where, the correct response probability to an item is P, g is an index indicating the latent class (g = 1, 2, …, G), and within class g examinee j's latent ability is θ_{jg} and for the class g item difficulty parameter is β_{ig}.

As a mixed distribution model, the MRM is a promising way to take into account qualitative individual differences without requiring the strong and necessary assumptions of the underlying non-mixed models (Rost, 1990).

In this study it is aimed to apply MRM to the TIMSS 2015 math test for determining seperate latent classes regarding math success among 8th grade students that vary in their item level patterns, if any. Thus, it can be examined whether the latent classes imply different response patterns, and also the properties of latent classes and the properties of items changing according to these latent classes. Besides these, the applications of MRM in scale validation by identifying qualitatively different latent classes are can be demonstrated. As stated Baghaei and Carstensen (2013), the presence of latent classes leads problems in interpretation and generalization of test scores and therefore it is a problem for test validity. So, the findings of this study can be useful for researchers who need to the interpretation math scores for revising or developing interested construct theories.

METHOD

Participants and Instrument

About half of the items in TIMSS 2015 are multiple-choice, and half are constructed items having long/short answers. In both classes (4th and 8th grades), science and mathematics items consist of 28 blocks. Of these blocks, 14 are science, and 14 are math blocks. These blocks were distributed to 14 test booklets in four blocks, two of which were science and two were mathematics. One of the two blocks in science and mathematics is expected between two booklets for test equating between forms (MoNE, 2016). For the study, the first booklet was chosen to apply the MRM.

The study sample consisted of 435 Turkish 8th-grade students (207 girls, 228 boys) answering booklet one item of Mathematics test in TIMSS 2015.

The booklet test includes 35 items scored dichotomously and partially credited. For using MRM in the analysis, items that scored partially credited were recoded dichotomously as 1-0. After recording these items, the unidimensionality assumption was checked with confirmatory factor analysis using diagonal weighted least square estimation in LISREL (ver. 8.80). The items having factor loadings of less than 0.30 and non-significant t-values were excluded from the test. It was seen that the unidimensional model fit the data over the remaining 28 items.

Data Analysis

The pattern of responses to 28 math items was analyzed using WINMIRA (von Davier, 2001). The number of latent classes was determined by an exploratory MRM analysis. Then, the parsimonious model that best fits the data was chosen. For this purpose, one, two, three and four latent class MRM analyzes were performed on the data and three information indices were compared to select the most appropriate model: “Akaike’s information criterion (AIC; Akaike, 1974), the Bayesian information criterion (BIC; Schwarz, 1978), and the consistent AIC (CAIC; Bozdogan, 1987). These criteria are computed as follows:

$$AIC = - 2 \log L + 2 p$$

$$BIC = - 2 \log L + p (\log N)$$

$$CAIC = - 2 \log L + p (\log N + 1)$$

where L is the likelihood, N is sample size and p is the number of estimated parameters in the model. AIC is not asymptotically consistent as sample size is not used in its calculation. BIC and CAIC select the models with fewer parameters compared to AIC. Models which have the lowest information criteria are selected in this study."

In addition to these information criteria, Q-index giving information about item fit was also estimated in WINMIRA. The item Q−index (Rost and von Davier, 1994) is calculated depending on Rasch model's parameter separability and conditional inference features (von Davier, 2001). The item−Q is estimated according to the log−likelihood of the observed item pattern. The Q index ranges from 0 to 1, a value of 0 indicates perfect fit, 0.50 indicates no relationship between item-measured features, and a value of 1 indicates negative discrimination. The standardized Q index, Zq, is the form of Q with a mean of 0 and a variance of unity. The widely used ±1.96 limit of the 95% confidence interval is applicable to the standardized Q index (Baghaei & Carstensen, 2013).

FINDINGS

Unidimensionality for the Test

Confirmatory Factor Analysis (CFA) was conducted to test the unidimensionality of the mathematics achievement test consisting of 35 items in booklet 1. According to Büyüköztürk (2018), 0.30 is sufficient as the lower limit for the factor loading values to be valid. Therefore, items with factor loadings below 0.30 in CFA were hierarchically excluded from the scale and analyzes were repeated. Seven items with factor loadings below 0.30 were excluded from the analysis. Accordingly, the standardized factor loading values for the items in the model obtained from CFA is more significant than 0.30. In the evaluation of the unidimensionality, in addition to the factor loadings, Comparative Fit Index (CFI), Normed Fit Index (NFI), Nonnormed Fit Index (NNFI) and Root Mean Square Error Index (RMSEA) compliance index values were calculated. For CFI, NFI and NNFI indices, 0.90 indicates acceptable fit, and 0.95 indicates perfect fit (Bentler & Bonett, 1980; Hooper, Coughlan & Mullen, 2008). For RMSEA, 0.08 is considered acceptable, and 0.05 is an excellent fit (Cheung & Rensvold, 2002; Marcoulides & Yuan, 2016). When fit indices for this model were examined, it was seen that CFI = 0.994, NFI = 0.982, NNFI = 0.993 and RMSEA = 0.033. When the perfect and acceptable fit criteria for fit indices are considered, it was seen that the fit indices obtained from CFA were sufficient. In the last case, the model fit data for unidimensionality was achieved.

Model Selection - Number of Latent Classes

Information criteria values for models with one, two, three and four classes required to identify to the convenient number of latent classes are given in Table 1.

Table 1. MRM Model fit Indices

Model	AIC	BIC	CAIC
One Class	12978.33	13096.52	13125.52
Two Classes	12767.46	13007.90	13066.90
Three Classes	12653.14	13015.84	13104.84
Four Classes	12659.94	13144.90	13263.90

When the AIC, BIC and CAIC values given in Table 1 for the four models were examined, it was seen the smallest BIC and CAIC values belong to a two-class MRM. However, the three-class MRM has the smallest AIC information criterion. It was decided that the data fit to the two-class MRM which is having the smallest values the BIC and CAIC information criteria which are more often recommended in the researches (Read & Cressie, 1988; Rost, 1996). For the best fitting model, latent class proportions and mean assignment probabilities were given in Table 2.

Table 2. *Two-Class MRM Mean Assignment Probabilities*

Latent Class	Latent Class proportions	Mean assignment probability	
		Class 1	Class 2
Class1	0,598	0,952	0,048
Class2	0,402	0,034	0,966

From Table 2, we see that latent class1, latent class2 proportions and mean assignment probability. Mean assignment probabilities of the students in latent class1 and latent class2 are 95.2 and 96.6 percent respectively. It can also be said that the two-class MRM has high mean assignment probabilities for both classes. According to the table 2, 260 (%59.8) of the students are in class 1 and 175 (%40.2) are in class 2.

The order of some items according to their difficulty parameters is presented in Figure 1. Specifically, the most difficult items, the easiest items and items with varying degrees of difficulties or between classes 1 and 2 are given in Figure 1.

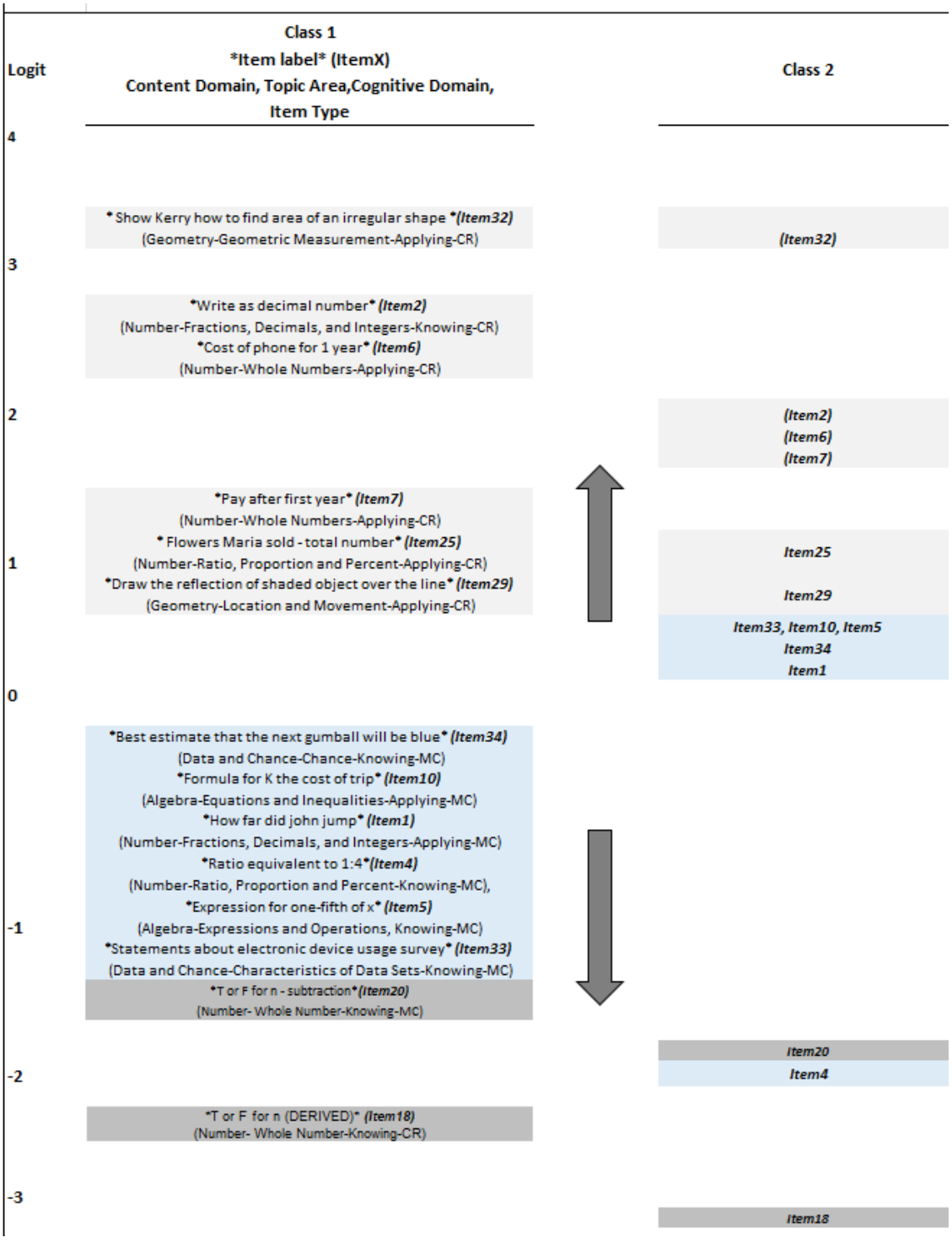

Figure 1. Ordering items according to their difficulty parameters in each latent class. Note: MC: Multiple Choice, CR: Constructed Reponses.

From Figure 1, items 32, 2, 6, 7, 25 and 29 for class 1 and class 2, respectively, were found similar and relatively more difficult than other items. While the content domains and topic areas of these items can be different, cognitive domains and item types are the same for all. The cognitive domains of the items (except item 2) are “Applying”, the item types are “Constructed Response”.

Specially, the most difficult item in both classes is item 32. This item "Show Kerry how to find area of an irregular shape" is geometry content domain that requires geometry measurement.

Items 34, 10, 1, 4, 5 and 33 were identified as relatively easy items in both classes. This situation can be interpreted in more detail as follows; Items 34, 10, 1, 5 and 33 were found to be more difficult in latent class2 than in latent class1, whereas item 4 was easier in latent class 1. When the cognitive domains and item types of these items differ in their difficulty levels in both classes, cognitive domains of all items except the 10th and 1st items are "Knowing", all of items are "Multiple Choice".

The items that are easier for students in both classes compared to other items are 20th and 18th items. The content domains of these items are "Number", cognitive domains are "Knowing".

Figure 2 shows item difficulty parameters for two latent classes. Horizontal axis shows the 28 items from 1 to 35, while the vertical axis shows the logit difficulty scale.

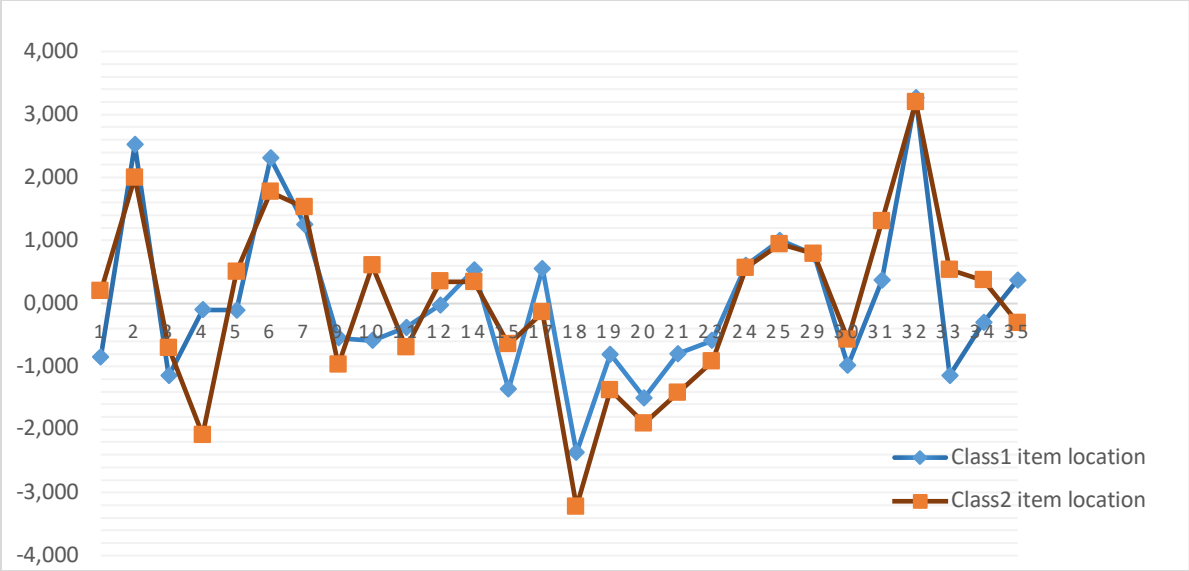

Figure 2. Item difficulty parameters for class 1 and class 2

When the figure is examined, it is seen that the some of items in the two latent classes have different difficulty parameters. The majority of the items (2, 4, 6, 9, 11, 14, items between 17-21, 23, 24, 25, 29, 32 and 35) are found easier in latent class2 than in latent class1. The others were easier for latent class 1. In this case, we can say that the test is easier for latent class2 or more difficult for latent class1. Especially, the difficulty parameters of item14, item24, item25, item29 and item32 between classes are very close to each other, that is, they have similar difficulty levels in both latent classes. While content areas of these items are "Geometry" (item 14, item 29 and item 32) and "Number" (item 24 and item 25), their cognitive domains are "Reasoning" (item 14) and "Applying" (item 24, item 25, item 29 and item 32).

The Relationships Between Latent Classes

In order to determine whether the latent classes obtained as a result of MRM differ significantly or not, it was performed for independent samples t-test. The t-test results of mathematics test scores according to the classes are given in Table 3.

Table 3. Mathematic test scores by class t-test results

Class	N	$\bar{X}$	Std. Deviation	df	t	p	η^2
Class 1	260	-1,255	0,58	433	30,406	0,00	0,681
Class 2	175	1,184	1,084				

As shown in from Table 3, it was seen that math scores differed significantly for latent classes, t (433) = 30,406, p <.01. In other words, the mean of class2 (M = 1.184, SD = 1.084) is significantly higher than the mean score of class1 (M = -1.255, SD = 0.58). When the explained variance η^2 value is examined, it is seen that 68% of the total mathematics achievement variance which is the dependent variable, is explained by the difference between latent classes. For η^2, also called effect size, the value calculated in this study is interpreted as a large effect size (0.01 low, 0.6 medium, and 0.14 high).

This finding is consistent with the findings about item parameters, such that that the majority of the items are found easier in latent class 2. So, it can be concluded that the second latent class is more successful in mathematics than the first class.

Item Fit for Each Latent Class

Lastly, in addition to model data fit, item level fit was examined with Q index (Rost &Davier, 1994). Item fit statistics were determined by the Q index and Z_q are given Table 4.

Table 4. Item fit statistics in two latent Classes

	Class 1		Class 2	
Item Label	Q-index (Zq)	p(X>Zq)	Q-index (Zq)	p(X>Zq)
I1	0,352 (0,220)	0,413	0,235 (0,595)	0,276
I2	0,339 (-0,078)	0,531	0,221 (-0,139)	0,555
I3	0,336 (0,154)	0,438	0,233 (0,414)	0,339
I4	0,329 (-0,100)	0,539	0,293 (0,103)	0,459
I5	0,346 (0,068)	0,472	0,175 (-0,413)	0,660
I6	0,216 (-0,735)	0,768	0,258 (0,497)	0,310
I7	0,273 (-0,425)	0,664	0,218 (0,036)	0,486
I9	0,361 (0,255)	0,399	0,196 (-0,210)	0,583
I10	0,393 (0,680)	0,248	0,180 (-0,260)	0,603
I11	0,332 (0,021)	0,491	0,192 (-0,158)	0,563
I13	0,278 (-0,452)	0,674	0,209 (0,280)	0,390
I14	0,325 (-0,423)	0,663	0,247 (0,520)	0,301
I15	0,291 (-0,456)	0,675	0,224 (0,220)	0,413
I17	0,226 (-1,133)	0,871	0,222 (0,218)	0,414
I18	0,352 (0,323)	0,373	0,199 (-0,083)	0,533
I19	0,369 (0,456)	0,324	0,255 (0,283)	0,389
I20	0,343 (0,254)	0,399	0,182 (-0,239)	0,594
I21	0,437 (1,128)	0,129	0,205 (-0,145)	0,558
I23	0,302 (-0,174)	0,569	0,264 (0,655)	0,256
I24	0,204 (-1,260)	0,896	0,184 (-0,264)	0,604
I25	0,231 (-0,886)	0,812	0,131 (-1,172)	0,879
I29	0,242 (-0,977)	0,835	0,195 (-0,191)	0,576
I30	0,273 (-0,504)	0,692	0,219 (0,281)	0,390
I31	0,411 (0,632)	0,263	0,144 (-1,057)	0,855
I32	0,286 (-0,183)	0,572	0,254 (-0,060)	0,524
I33	0,360 (0,324)	0,372	0,169 (-0,401)	0,656
I34	0,338 (-0,143)	0,556	0,290 (1,335)	0,091
I35	0,451 (1,061)	0,144	0,182 (-0,331)	0,630

As shown in Table 4, Q index values for latent class 1 vary between 0,204 and 0,451. For latent class2, the Q index values range between 0,131 and 0,293. In addition to the Q index values, when the Zq values are examined, it can be said that all items in both classes have non-significant Zq values, that is, in item level model fit to data for both latent classes.

DISCUSSION AND CONCLUSION

This study used MRM to examine differences in TIMSS math success scores. As a result of the analysis, two-class MRM with sizes 0.598 and 0.402 has a better model fit to the data than a standard one-class model for the math test. In general, students in class 2 are more successful than those in class 1 according to their test scores obtained with MRM. Item-level model data fit was assessed via Q and Zq indices, and it was concluded item level model fit was provided. In the study of De Ayala and Santiago (2017), students' mathematical abilities were tested with MixIRT models. Similar to the results of this study, the best model fit is a two-latent class mixture one-parameter model. In addition, some items were found more accessible for one latent class, others harder.

The classes identified in the current study were similar in size but had a different patterns of math achievement. The complex items for both groups are in the “applying” cognitive domain and have constructed response types. The content domain of the more accessible items for students in both classes is “number,” and the cognitive domain regarding these items is “knowing.” When TIMSS 2015 8th grade mathematics questions are examined in terms of a cognitive level conceptual framework, 33% of the questions are at the level of “knowing,” 45% are at the level of “applying,” and 22% are at the level of “reasoning” (Mullis et al., 2016). In other words, it can be said that most of the mathematics questions are based on students' use of problem-solving instead of basic definitions and simple calculations.

The fact that the items that are difficult for both groups are at the applying level can be a valuable finding for the teachers' practices in the classroom and giving homework in the school setting. Teachers can focus on practical situations and real-life problems, which enable students to apply what they know about mathematics in the classroom learning environment. Kazmierzak (1994) stated in his study that practices requiring high-level thinking skills are rarely given by teachers and emphasized that assigning similar problems solved in the course as the practice is ineffective. Incikapı et al. (2016) conducted a study to analyze the cognitive domains of the acquisitions identified in the middle school mathematics curriculum according to grade level (grades five through eight) and content domains with document analysis. It was examined according to the cognitive domains and their sub-domains identified in TIMSS 2015 mathematics framework. As the results of the study, although the distribution of the cognitive domains of the acquisitions differs by the grade levels, for knowing, applying and reasoning, the total number of acquisitions in cognitive areas are very close. In this case, it can be interpreted that the students' difficulties at the applying level might be caused by classroom practices or assessment techniques rather than the curriculum.

In addition, it is necessary to ensure that students are more familiar with the questions they have to construct the answer rather than the multiple-choice ones. For this purpose, measurement and evaluation techniques should be used in which students can construct the answers themselves when they encounter a problem.

There are two sub-groups related to mathematics achievement in the study group because the difficulty levels of the items differ between the classes. Again, there is a significant difference in achievement between the classes. So, two subtypes of strategies exist among students.

Mislevy and Huang (2007) addressed that individuals belonging to different latent classes may be due to different education systems and curricula or the application of different strategies to the solution. A thorough review of item contents can provide beneficial information about the qualitative differences among the examinees. However, in the current study, having access to the actual math items is a limitation. In TIMSS data, only some test specifications like cognitive domains and content areas were provided for the researchers for other minor analyses. That is why deeply examining the contents of items with differing difficulty estimates across latent classes was impossible. Such assessments can provide a more profound comprehension of the development and process involved in math achievement.

Baghaei and Carstensen (2013) stated that MRM has a wide application in developmental psychology. It can be investigated if there are different types of learners with different patterns of learning and if those learner types may be associated with external factors such as age, gender, motivation, first language, etc. For example, as stated by Glück and Spiel (2007, p. 292), considering age differences in item response patterns may be necessary for test development. The Participants at different age levels can reveal unusual processes of change that can contribute to an understanding of development. For this reason, external factors that cause different patterns with MRM applications can be investigated, both confirmatory and exploratory (Baghaei & Carstensen, 2013).

RECOMMENDATIONS

In future studies, in cases where the items are created by the researchers or provided access to full forms of the items, more detailed analyzes can be achieved and more detailed comments can be made about the students' solution strategies. MRM analyses of math tests with well-organized items may answer questions about the nature of math achievement and subgroups of math ability and the relation between these subgroups and other variables.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, author-ship, and/or publication of this article.

Statements of publication ethics

I/We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully.

Researchers' contribution rate

The study was conducted and reported with equal collaboration of the researchers.

Ethics Committee Approval Information

In this study, TIMSS 2015 8th grade mathematics data, which is open to everyone, was used. Said data was downloaded from https://timssandpirls.bc.edu/timss2015/international-database/ web page. That is, data which is used in the study dated before year of 2020 and open access. Therefore, no ethics committee approval was obtained.

REFERENCES

Baghaei, P., & Carstensen, C. H. (2013). Fitting the Mixed Rasch Model to a Reading Comprehension Test: Identifying Reader Types. *Practical Assessment, Research & Evaluation*, *18*.

Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980; 88(3):588-606.

Büyüköztürk Ş. (2018). *Sosyal bilimler için veri analizi el kitabı [Manual of data analysis for l sciences].* Ankara: Pegem Akademi.

Cheung GW, Rensvold RB. Evaluating goodnessof-fit indexes for testing measurement invariance. *Structural Equation Modeling 2002; 9*(2):233-255.

Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, *42*(2), 133-148.

De Ayala, R. J. & Santiago, S. Y. (2017). An introduction to mixture item response theory models. *Journal of School Psychology,* 60, 25-40. doi: 10.1016/j.jsp.2016.01.002.

Embretson, S. E. (2007). Mixed Rasch models for measurement in cognitive psychology. In M. Von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: extensions and applications* (pp. 235-253). New York: Springer Verlag.

Glück, J., & Spiel, C. (2007). Studying development via Item Response Model: A wide range of potential uses. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 281-292). New York: Springer Verlag.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.

Hooper D, Coughlan J., Mullen MR. (2018). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods 2008; 6*(1):53-60.

İncikabı, L., Ayanoğlu, P., Aliustaoğlu, F., Tekin, N., & Mercimek, O. (2016). Ortaokul matematik dersi öğretim programı kazanımlarının TIMSS bilişsel alanlarına göre değerlendirilmesi [An Evaluation of Middle School Mathematics Teaching Programs Based on TIMSS Cognitive Domains.] *İlköğretim Online*, *15*(4).

Kazmierzak K. S. (1994). *Current wisdom on homework and the effectiveness of a homevvork checking system. (*Report No: E/S 591) Indiana: Indiana University at South Bend.

Li, T., Jiao, H., & Macready, G. B. (2016). Different approaches to covariate inclusion in the mixture Rasch model. *Educational and psychological measurement*, *76*(5), 848-872.

Marcoulides KM, Yuan K. (2016). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models.

Ministry of National Education (2016). TIMSS 2015 national mathematics and science preliminary report for 4th and 8th grades. Ankara, Turkey: Ministry of National Education General Directorate of Measurement, Evaluation and Examination Services.

Mislevy, R., & Huang, C.-W. (2007). Measurement models as narrative structures. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 15-35). New York: Springer Verlag.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195 -215.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. Retrieved 22.12.2019, from http://timssandpirls.bc.edu/timss2015/internationalresults/

Ölmez, İ. B., & Cohen, A. S. (2018). A mixture partial credit analysis of math anxiety. *International Journal of Assessment Tools in Education*, *5*(4), 611-630.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, 1960. (Expanded edition, Chicago: The university of Chicago Press, 1980).

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271 – 282.

Rost, J., Carstensen, C., & von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. Rost, & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 324-332). Munster, Germany: Waxmann.

Rost, J. & Davier, von, M. (1994). A conditional Item Fit Index for Rasch Models, *Applied Psychological Measurement.*

Tavsancil, E., & Yalcin, S. (2015). A determination of Turkish student's achievement using hierarchical linear models in trends in international mathematics-science study (TIMSS) 2011. *The Anthropologist*, *22*(2), 390-396.

Von Davier, M. (2001). WINMIRA [Computer Software]. St. Paul, MN: Assessment Systems Corporation.