Necmettin Erbakan Üniversitesi
Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF)
Journal of Ahmet Keleşoğlu Education Faculty
DOI: 10.38151/akef.72

Cilt: 5 Sayı:2 Yıl: 2023 | Araştırma Makalesi / Research Article | E-ISSN: 2687- 1750

An Analysis of the DIF Sources of ABIDE Mathematics Self-Efficacy Scale by means of a Latent Class Approach*

Fuat ELKONCA[1] **İsmail KARAKAYA**[2]

[1]Dr. Öğr. Üyesi, Muş Alparslan Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Muş, Türkiye f.elkonca@alparslan.edu.tr (Corresponding Author)

[2]Prof. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara, Türkiye ikarakaya@gazi.edu.tr

Article Info

Article History
Received: 11.08.2023
Accepted: 20.09.2023
Published: 30.09.2023

Keywords:
Laten Class Analysis
Mixture Logistic Regression
Third Generation DIF
ABIDE
Self-efficacy

ABSTRACT

This study aims to identify the sources of Differential Item Functioning (DIF) using the Mixture Ordinal Logistic Regression (Mixture OLR) method, a contemporary approach for detecting DIF. To analyze mathematics self-efficacy, data from a scale comprising 9 items were obtained from 5000 8th-grade students as part of the ABIDE-2016 project. The study compared the presence and extent of DIF by gender using two methods and examined the sources of DIF for items displaying DIF with Mixture OLR. The OLR analysis revealed that five items exhibited DIF at level A, but no DIF was observed with Mixture OLR. Furthermore, it was found that the magnitude of DIF (B) for an item showing DIF at level A changed due to Mixture OLR. The results indicate that the homogeneity of the data affects both the number of items displaying DIF and the magnitude of DIF. Three items did not exhibit significant DIF according to both methods. One significant finding in the study highlights the moderating effect of latent class on item 8, where DIF was observed. However, the source of DIF was not related to gender but rather stemmed from different ecological variables. An analysis of latent class characteristics revealed that students with significant DIF effects had lower absenteeism and fewer siblings. Additionally, students in this class had greater access to books at home and participated in more out-of-school mathematics courses. Surprisingly, these students were found to engage less in social activities. Various factors can influence how students respond to test items, potentially leading to DIF. These factors may include cultural background, gender, social environment, school, teacher, family interest/attitude toward the child, and home climate. Therefore, when developing and administering tests, it is crucial to test for data homogeneity and consider the impact of these variables, in addition to gender, to identify any sources of DIF in test items.

ABİDE Matematik Özyeterlik Ölçeği DMF Kaynaklarının Gizil Sınıf Yaklaşımıyla İncelenmesi

Makale Bilgileri

Makale Geçmişi
Geliş: 11.08.2023
Kabul: 20.09.2023
Yayın: 30.09.2023

Anahtar Kelimeler:
Gizil Sınıf Analizi
Karma Lojistik Regresyon
Üçüncü Nesil DMF
ABİDE
Özyeterlik

ÖZ

Bu araştırmanın amacı, Değişen Madde Fonksiyonu (DMF) belirlemede güncel bir yaklaşım olan Karma Ordinal Lojistik Regresyon (Karma OLR) yöntemini kullanarak DMF kaynaklarının belirlenmesidir. Bu amaçla 9 maddeden oluşan matematik özyeterlik ölçeğinden elde edilen veriler kullanılmıştır. Araştırma kapsamında Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE-2016) projesi kapsamındaki 5000 8.sınıf öğrencisinden elde edilen veriler kullanılmıştır. Araştırmada cinsiyete göre DMF'nin varlığı ve büyüklüğü iki yöntem sonuçların açısınndan karşılaştırılmış ve DMF gösteren maddelerin DMF kaynakları Karma OLR ile incelenmiştir. OLR kullanılarak yapılan analizlere göre 5 maddenin A düzeyinde DMF gösterdiği sonucuna ulaşılmıştır. Aynı maddeler Karma OLR ile incelendiğinde ise maddelerde DMF görülmemiştir. Bu sonuca ek olarak A düzeyinde DMF gösteren bir maddenin Karma OLR sonucunda DMF büyüklüğünün (B) değiştiği görülmüştür. Elde edilen sonuçlar homojenliğin DMF gösteren madde sayısı ve DMF'nin büyüklüğünü etkilediğini göstermektedir. Üç madde de her iki yönteme göre de DMF anlamlı çıkmamıştır. Araştırmada elde edilen diğer bir önemli sonuç ise gizil sınıfın moderetör etkisinin anlamlı çıktığı bir maddede (madde 8) DMF kaynağının cinsiyetten ziyade farklı ekolojik değişkenlerden kaynaklandığıdır. Gizil sınıf özellikleri incelendiğinde, özellikle DMF etkisinin anlamlı olduğu sınıftaki öğrencilerin devamsızlık durumu ve kardeş sayılarının daha düşük olduğu görülmüştür. Bununla beraber bu sınıfta yer alan öğrencilerin evde var olan kitap sayıları ve matematik için okul harici kurs alma durumlarının yüksek olduğu sonucuna ulaşılmıştır. Dikkat çekici sonuç ise bu öğrencilerin sosyal faaliyetlere daha az katılmalarıdır. Öğrencilerin maddelere verdikleri tepkileri etkileyen bu değişkenler DMF kaynağı olarak yorumlanabilir. Çünkü maddeye verilen tepkiler sadece kültür ya da cinsiyet gibi özelliklerle ilişkili olmayabilir. Bu gibi karakteristik özelliklere ek olarak, madde tipi, sosyal çevre, okul, öğretmen, çocuğa yönelik aile ilgi/tutum durumu ve ev iklimi gibi birçok değişkenden etkilenebilir. Buradan haraketle özellikle test uygulama ve geliştirme süreçlerinde veri yapısının homejenliği test edilmeli ve DMF'li maddelerin kaynakları incelenirken cinsiyet gibi karakteristik özelliklerle beraber madde tepki örüntüsünü etkileyebilecek değişkenler de dikkate alınmalıdır.
.

*This article is derived from part of his doctoral dissertation titled "An analysis of the DIF sources of ABIDE self-efficacy scale by means of a latent class approach"

Atıf/Citation: Elkonca, F. & Karakaya, İ. (2023). An Analysis of the DIF Sources of ABIDE Mathematics Self-Efficacy Scale by means of a Latent Class Approach. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF) Dergisi*, 5(2), 579-594.

"This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)"

INTRODUCTION

In Türkiye, many educational decisions and training process are based on the results obtained from national and international exams. In our country, measurement and evaluation studies are carried out periodically at national and international level. At the international level, exams are implemented within the scope of projects, such as "Program for International Student Assessment" (PISA), "Trends in International Mathematics and Science Study" (TIMSS) and "Progress in International Reading Literacy Study" (PIRLS). Especially as an alternative to the PISA exam, the Monitoring and Evaluation of Academic Skills (ABIDE) exam is implemented at the national level in Türkiye (MEB, 2016a; 2016b; 2016c). Within the scope of the ABIDE project, which has been applied to eighth-grade students since 2016, cognitive characteristics and affective characteristics such as self-efficacy and attitude are measured in Turkish mathematics, science, and social studies courses. The ABIDE project aims to determine the extent to which students have cognitive skills in these subject areas and to reveal the affective, family, and school characteristics that affect their achievement. The accuracy of the decisions made based on the measurement results of such large-scale exams is closely related to the reliability and validity of the measurement results obtained through the applications (Yalçın & Tavşancıl, 2015). Indeed, Cronbach (1984) defined validity as the process of gathering evidence to support inferences that can be drawn from test scores. The more evidence collected, the more information is obtained about the validity of the test (Kelecioğlu & Şahin, 2014). Therefore, researchers should focus on collecting evidence of validity in all processes, from the development of a test or scale, the response processes of individuals, the administration of the test or scale, and the interpretation of scores, and increasing the quality of this evidence.

One of the factors that may affect validity is that test items show different psychometric properties among individuals with different cultural, demographic, social, and linguistic background experiences (Gomez-Benito et al., 2018). This concept, considered bias in the literature, is explained as the situation in which a test item provides an unfair benefit to one group compared to other groups (Clauser & Mazor, 1998). Item bias is the difference in the probability of answering the item correctly for two groups at the same ability level due to the characteristics of the test items or test conditions unsuitable for the test (Zumbo, 1999). The concepts of item effect and Differential Item Functioning (DIF) should also be mentioned for a better understanding of item bias. If the differences between the probabilities of answering the item correctly of participants in different groups are real group differences in terms of the ability intended to be measured by the item, this is defined as an item effect. DIF, on the other hand, occurs when participants in different groups differ in their probability of answering the item correctly after matching in terms of the ability intended to be measured by the item (Camilli & Shepard, 1994; Clauser & Mazor, 1998; Zumbo, 1999). In observed group DIF approaches, many methods are used to determine whether items show DIF (e.g., Mantel Haenszel, Chi-square, SIBTEST, Logistic regression, IRT models, etc.). In the literature, DIF techniques can be classified according to whether they are parametric or not, whether the matching variable is observed or latent, whether the item scores are dichotomous or multi-categorical, and whether they can detect uniform or non-uniform DIF (Wiberg, 2007).

The above-mentioned manifest group DIF analyses are based on the rather strong assumption that all members of an observed group (e.g., gender) use the same strategies or problem-solving techniques and have the same experience with item content. Therefore, in the traditional DIF approach, although the probability of responding correctly to an item depends on the respondent's ability, all members of an observed group view the item similarly. An example of this model is that an item that shows DIF for males is disadvantageous for all males. Samuelsen (2008) found that removing the item displaying DIF would increase the scores of all men. However, it is crucial to note that observed group variables, such as gender, may not always represent homogeneous groups, as emphasized by De Ayala at al. (2002). Furthermore, observed group DIF analyses have limitations, such as weak correlation between gender and DIF, as noted by Kang & Cohen (2003) and Samuelsen (2005).

Due to the limitations mentioned above of the observed group DIF approaches, latent class DIF

approaches are frequently used in recent studies on DIF (e.g., Cho at al., 2016; Cohen & Bolt, 2005; DeMars & Lau, 2011; Frick at al., 2015; Lee Webb et al., 2008; Oliveri et al., 2013; Oliveri at al., 2016; Samuelsen, 2008; Zumbo et al., 2015). In the latent class approach, individuals are assigned to one of the latent classes based on their responses. Individuals in each latent class are seen as having similar or homogeneous response patterns (Mislevy at al., 2002; Oliveri et al., 2013). In contrast to observed group DIF approaches where all individuals in certain groups, such as gender, are assumed to have homogeneous response patterns, different proportions of girls or boys can be assigned to a class when using a latent class DIF approach (Oliveri et al., 2013). Such a heterogeneous data structure implies one or more latent classes and a different distribution within each latent class. Therefore, this approach assumes qualitative differences between participants in responding to the test item and that these differences are due to latent classes rather than observed groups (Samuelsen, 2008). For this reason, it is important to use latent class regression models, which divide a heterogeneous data set into homogeneous subgroups (latent class) and make more accurate and unbiased predictions by making a separate prediction for each group (Kayri, 2006), in DMF-related research.

Another significant advantage of using the latent class DIF approach is that it is helpful to access the sources of DIF (Lee Webb et al., 2008; Oliveri et al., 2013; Samuelsen, 2008; Zumbo, 2007). Because the latent class DIF approach focuses on latent class memberships rather than observed groups, covariates (teacher, student, school, test conditions, etc.) can be included in the model as predictors of latent class memberships. In this way, more detailed information about why DIF may have occurred can be obtained (Samuelsen, 2008).

Latent class DIF methods are considered one of the third generation DIF approaches. Accordingly, “while the first generation DIF approach focused more on the psychometric problems of test bias and some essential concepts, the second generation DIF approach focused on the prominence of the term DIF, the multidimensional structure of tests, the development of appropriate statistical tests and the comparison of their effectiveness” The third generation DIF approach focuses on the multidimensional nature of the tests and the development of appropriate statistical tests and comparisons of their effectiveness. In other words, in the third generation DIF approach, DIF is considered as follows; DIF is caused by some characteristics of the test item and/or the test situation that are unrelated to the relevant ability and, therefore, not related to the test's purpose. By adding and emphasizing the test situation as a possible cause of DIF, the perspective on DIF is greatly expanded and goes beyond the test construct (Zumbo et al., 2015). This approach allows emphasizing sociological, structural, social, and contextual variables, as well as psychological and cognitive factors that cause individuals' item response patterns to change, as explanatory sources of item response and thus of DIF (Zumbo & Gelin, 2005).

There are various methods in which DIF sources are investigated using latent class analysis. These methods are a Mixture of Rasch models based on IRT and Latent Class Logistic Regression models (Cohen & Bolt, 2005; De Ayala et al., 2002; Rost, 1990; Samuelsen, 2005). Mixture models based on IRT do not provide an overall test or measure of DIF for each item but only allow some parameters to vary across two or more classes. In Latent Class LR models, a separate regression model is constructed for each item, similar to the standard LR method. However, if the data set is not homogeneous, that is, if more than one latent class is significant, the regression model is tested separately for each category. In this way, the regression coefficients obtained for an item in different classes can be compared. In addition, the Mixture LR method provides alternative methods for many different item formats (binary, ordinal, nominal). If the dependent variable is at the ordinal scale level, the Mixture Ordinal LR (Mixture OLR) method is used (Vermunt & Magidson, 2005). The mixture LR method includes both the parameters of observed variables independent of class and the parameters obtained by testing these variables in different latent classes. Since this method allows the regression model established for an item to be tested in different latent classes, it can also test the moderating effect of the latent class on DIF. In other words, the DIF effect may vary across classes. By adding covariates to the class memberships obtained in this way, DIF sources can be examined in more detail. In these aspects, Mixture LR is a very suitable method for the third generation DIF approach (Zumbo et al., 2015).

In this study, whether the items in the self-efficacy scales of the ABIDE exam administered by Ministry of National Education (MoNE) at the national level show DIF according to gender, and the DIF sources of the items showing DIF were examined with the Mixed OLR method, which is one of the third generation DIF approaches. Many variables affect students' item response patterns. For example, most studies examining DIF by gender have focused on item characteristics such as item content and item format that may affect students' test performance. However, many variables, such as individual characteristics, school characteristics, class size, socioeconomic status, teaching practices, and parenting styles, can also affect students' item responses and few studies (e.g., Zumbo & Gelin, 2005) have considered these factors. Therefore, this study aims to examine the sources of students' differentiated item responses on the items in the ABIDE self-efficacy scales in a detailed and holistic manner using the third generation DIF approach and the ecological model. This study examines whether the items in the Mathematics Self-Efficacy scale of the 2016 ABIDE exam show DIF according to gender and the sources of DIF with the latent class DIF approach. In line with this primary purpose, the following questions were sought to be answered:

1. Are there any items in the 2016 ABIDE Mathematics Self-Efficacy scale that show DIF according to gender in the analyses conducted with the OLR method?

   2016 ABIDE Mathematics Self-Efficacy Scale based on each item:

1.1. Does it contain more than one latent class regarding the measured characteristic?

   If there is more than one latent class:

1.1.1. Do the parameter estimates (DIF by gender) made according to the mixture OLR method differ in latent classes?

1.1.2. Are there any variable or variables (ecological variables such as teacher, school, family, attitude, interest, etc.) that are significant predictors of latent class membership?

1.1.3. Does the presence and magnitude of DIF vary according to the method used (OLR and Mixture OLR)?

METHOD

Research Design

Within the scope of the research, whether nine items in the mathematics self-efficacy scale of the ABIDE test show DIF and the sources of DIF were examined with the Latent Class DIF approach. Since the relationships between the items in the mathematics self-efficacy scale used in the ABIDE exam will be examined in this study, the research is of correlational research type. “Correlational studies are studies in which the relationship between two or more variables is examined without intervening in these variables in any way” (Büyüköztürk at al., 2014).

Participants

The population of the ABIDE 2016 study consists of all eighth-grade students in Turkey. The MoNE selected the sample through a stratified sampling method among eighth-grade students regardless of school type. ABIDE 2016 application was realized with the participation of 35000 students selected among all eighth-grade students. In the study, data belonging to 5000 students randomly selected among 15000 students who took the same booklet in the ABIDE exam were used. In total, 800 students who did not respond to the tests of the courses of the variables and 40% of the self-efficacy scale, or all of any scale were excluded from the sample. As a result, 4191 individuals (2100 boys and 2091 girls) were used in the study. Multiple imputation method was used for missing data imputation. For this purpose, five different imputations were made and averaged for missing data in continuous and ordinal variables. For categorical variables, only one imputation was made.

Research Instruments and Processes

ABIDE mathematics self-efficacy scale and demographic information form were used within the scope of the research. There are 9 items in the applied form of the ABIDE mathematics self-efficacy scale. Confirmatory Factor Analysis (CFA) was performed to provide evidence for the validity of the measurement results obtained from this scale. The items in the scale and the factor loadings related to these items are given in Table 1.

Tablo 1. *Items in the Mathematics Self-Efficacy Scale and their Factor Loadings*

	Self-Efficacy Items	Factor Loadings
Item 1	I learn mathematic easily.	0.81*
Item 2	I can solve difficult mathematics questions.	0.84*
Item 3	Studying for a math exam takes a lot of time.	0.20*
Item 4	I am better at math than my classmates.	0.82*
Item 5	Mathematics lesson is no different for me from other lessons.	0.47*
Item 6	My teacher says I am good in Mathematics.	0.84*
Item 7	I ask my teacher about the things I don't understand in the math class.	0.63*
Item 8	I do not worry about failing the mathematics exams	0.55*
Item 9	My parents want me to get a high grade in Mathematics.	0.35*
Cronbach Alpha (α)		0,80

According to the mathematics self-efficacy scale DFA results, RMSEA value was .064, CFI, NFI, NNFI, AGFI values were between .90 and .98 and $\chi 2$ / df values were 32.75. The fact that "$\chi 2$ / df" ratio is less than 5, RMSEA and SRMR values are lower than .08, whereas CFI, NFI NNFI and AGFI values are higher than .90 indicate that the model fits the data well (Hu & Bentler, 1999; Tabachnick & Fidel, 2001). Although the goodness of fit values obtained as a result of the CFA analysis generally show that the model fits the data well, it is seen that the $\chi 2$ / sd values are quite high. The findings show that the model-data fit of the scales is not at a good level. The Cronbach Alpha (α) reliability coefficient was obtained as 0.80. This value shows that the reliability is at an acceptable level. The variables included in the demographic information form and included in the model as covariates are given in Table 5 below.

Data Analysis

Within the scope of the research, the Mixture Logistic Regression Method, one of the third generation DIF approach methods proposed by Zumbo (2007) and Zumbo et al. (2015), was used to identify the items showing DIF and the sources of DIF. DIF study with a Mixture LR method generally consists of four steps: In the first stage, traditional LR DIF analysis is performed. In the second stage, the number of latent classes is determined. After the number of latent classes is determined, DIF analysis is completed simultaneously and separately for each latent class using the Mixture LR method. If the moderating DIF effect of the latent class is significant, the significance of the coefficients obtained separately in each class is examined at this stage.

Within the scope of the research, the magnitude of the DIF obtained at the level of latent classes is calculated by making model comparisons in the class or classes where the DIF is effective. If the magnitude of the DIF was negligible (A), the third and fourth steps of the analysis did not proceed. While determining the number of latent classes, likelihood ratio (L2) test, AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), Npar (number of parameters), Classification Error (Class.Err.) and Explained variance (R2) values were taken into consideration. A detailed explanation of these indices is given in the theoretical framework. Although theoretical structure, interpretability, and simplicity are essential in model selection, in this study, the acceptance of the H0 hypothesis established by the likelihood ratio (L2) test, small values of BIC, AIC, classification error, and number of parameters, and high explained variance (R2) were considered as model selection criteria.

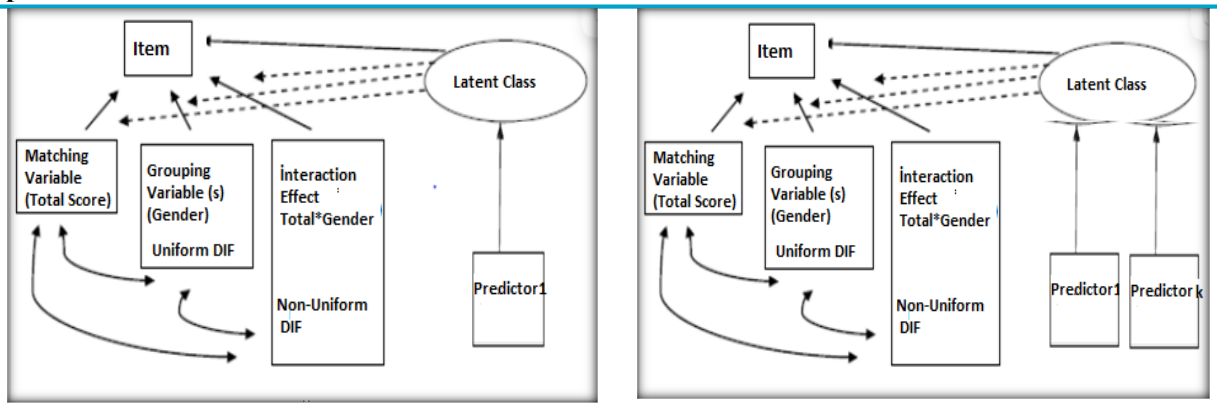

Figure 1. *Identification of latent class predictors by ecological variables* (Zumbo et al., 2015)

The 3rd and 4th stages are shaped according to the other stages. If the latent class has a moderating effect, then determining the variables predicting the latent class memberships will also provide information about the sources of DIF (Zumbo, 2007; Zumbo et al., 2015). Ordinal Logistic Regression (OLR) and Mixture Ordinal Logistic Regression Analysis (Mixture OLR) were used since the items in the mathematics self-efficacy scale used in the study were in a five-point Likert structure. In the fourth stage of the analysis, only one of the variables with a correlation value above 0.80 was included in the model since the significant variables were formed together. In addition, the interaction effect of the variables added to the model together in the fourth stage were also considered. Within the scope of the research, the analyses were conducted in the Latent Gold 5.0 program. A significance level of 0.05 was considered for all statistical tests.

Ethic

This article was found ethically appropriate with the decision number E_33941 of the scientific research and publication ethics committee of Gazi University of Applied Sciences on 04.03.2020.

RESULTS

In this section, the findings obtained in line with the aims of the study are presented. In this context, the items in the mathematics self-efficacy scale were analyzed separately. Standard OLR analysis was conducted in the first stage, and Mixture OLR analysis was conducted in the second stage. For the items where more than one latent class was significant, and the DIF magnitudes obtained were at the B and C level, the third and fourth stages of the analysis were started, and the DIF sources in the related items were examined. However, DIF sources were not examined for items with a negligible DIF effect in the latent classes, and this effect did not differ between the classes. In addition, in models where more than one latent class was not significant, the third and fourth stages of the analysis were not carried out. In this section, the findings of the analysis of the items in the ABIDE mathematics self-efficacy scale are presented. In this context, the findings of the 8th item, for which the 3rd and 4th stages of the Mixture OLR analysis can be applied, are presented as an example. Analyzes of other items are not presented here because they contain too many tables. The descriptive statistics of item 8 and the findings of the OLR analysis are given in Table 2.

Table 2. *Descriptive statistics and ordinal logistic regression DMF analysis for item* 8

Descriptive Statistics			**Logistic Regression DIF**				
Item	***Gender***	***X̄ (SS)***	***Variables***	***β***	***Wald***	***p***	**DIF Level**
I do not worry about failing math exams	Female	2.99 (1.51)	Total	0.10	732.58	.000*	
	Male	3.30 (1.43)	Gender	-0.41	7.74	.005*	A
	Total	3.15 (1.47)	Total*Gender	0.01	1.63	.200	

*p<.05

As is seen in Table 2, when the general average scores obtained from item 8 are considered, it is seen that male students scored higher than female students ($\bar{X}$ girl=2.99; $\bar{X}$ boy=3.30), and the general average score of the group was $\bar{X}$ = 3.15. Considering the OLR results, it is seen that while total score (β=0.10; Wald=732.58; p<.05) and gender (β=-0.41; Wald=7.74; p<.05) are significant on the scores given to item 8, the interaction effect of total*gender (β=0.01; Wald=1.63; p>.05) is not significant. The findings show that the uniform DIF is significant in the one-class model. Considering the β value estimated for gender, it is seen that DIF favors men. In the model comparisons, it is seen that the DIF level is A. In the second stage of the analysis, it was tested whether the responses to item 8 contained more than one latent class, in other words, whether the regression model based on item 8 was suitable for a single-class structure. The findings obtained are given in Table 3.

Table 3. *Results of latent class models for item 8*

Model	*LL*	*BIC(LL)*	*AIC(LL)*	*Npar*	*L²*	*df*	*p*	*Class.Err.*	*R²*
1st Grade	-5434.22	10926.83	10882.45	7.00	1294.86	285.00	.000*	0.00	0.44
2nd Grade	-5154.26	10433.63	10338.52	15.00	734.93	277.00	.000*	0.25	0.48
3rd Grade	-5072.55	10336.93	10191.10	23.00	571.51	269.00	.000*	0.28	0.74
4th Grade	-5044.82	10348.20	10151.64	31.00	516.05	261.00	.000*	0.31	0.93

*p<.05

As is seen in Table 3, the lowest BIC value is obtained for the three-class model (BIC=10336.93). It is seen that the BIC value and classification error increase after the three-class model. The findings indicated that the responses to item 8 are not homogeneous and contain three different latent classes. The regression coefficients estimated for the three-class model are given in Table 4.

Table 4. *Results for DIF and moderator DIF (mixture OLR) effect for the three-class model*

Predictors	***Grade1***		***Grade2***		***Grade3***					
	β	***Z***	***β***	***Z***	***β***	***Z***	***Wald***	***p***	***Wald(=)***	***p***
Total	0.56	12.43	0.14	8.71	0.07	2.32	306.57	.000*	70.73	.000*
Gender	-0.82	-1.34	-3.84	-4.88	-0.87	-1.04	29.78	.000*	10.58	.005*
Total*Gender	0.00	0.21	0.10	4.53	0.03	0.86	22.51	.000*	9.73	.008*

*p<.05

As is seen in Table 4, the total score effect (Wald=306.57; p<.05), gender effect (Wald=29.78; p<0.05), and interaction effect (Wald=22.51; p<.05) on item 8 in the three-class model were significant. In other words, both uniform DIF and non-uniform DIF were significant. Similar results were obtained in the Wald(=) test which was conducted to determine whether the coefficients obtained separately for each class differed between the classes. The regression coefficients obtained for the three grades differed significantly across grades for total score (Wald(=)=70.73; p<.05), gender (Wald(=)=10.58; p<.05) and total*gender (Wald(=)=9.73; p<.05). The findings showed that for item 8 in the mathematics self-efficacy scale, the DIF effect by gender in the single-class model and the moderating DIF effect of latent class is significant. In addition to these findings, the Z test was reported to determine whether the regression coefficients (β) estimated for each grade were significant only in the relevant grades. According to the findings, the interaction effect of gender (β=-3.84; Z<-1.96) and total*gender (β=0.10; Z>1.96) was significant only in 2nd grade. These effects were not significant in other grades. However, the β values obtained for the total*gender effect (βgrade1=0.00; βgrade2=0.10; βgrade3=0.03) are very close to zero, indicating that the effects are not significant. In other words, it can be said that the non-uniform DIF effect can be neglected. Based on these findings, it is seen that the DIF effect by gender is not effective in the whole group but only in the 2nd grade, and there is a DIF effect in favor of males. It was concluded that the magnitude of the DIF obtained in the model comparisons made only for class 2 was at the level of B (DIF magnitude). At this stage, the third and fourth stages of the analysis will be started. The latent class characteristics will be described by determining the covariates influential on the three-class latent structure, especially in the 2nd grade. In this way, the variables that cause the change in the response patterns of male and female students towards item 8, in other words, the variables that drive and do not

cause DIF, will be determined. At this stage, explanatory variables were first added to the model one by one and separately to determine the variables that were effective on the students' response patterns, especially in class 2. The findings obtained are given in Table 5.

Table 5. *Covariates added separately to the model in the mixed OLR DIF analysis for item 8*

Variable Types	*Explanatory Variables*		*Grade1*	*Grade2*	*Grade3*	*Wald*	*p*
Family-related variables	Number of siblings		0	-0.19	0.06	16.40	.000*
	Father's education level		0	0.21	0.07	27.21	.000*
	Mother's education level		0	0.16	0.01	15.77	.000*
	Monthly income		0	0.09	0.00	3.18	.200
	Family interest level		0	0.02	-0.05	7.48	.024*
	Family bullying level		0	-0.01	-0.03	1.15	.560
Variables related to Personal Characteristics - Individual Differences	Abide Math Achievement Score		0	0.01	0.00	106.29	.000*
	Level of participation in social activities		0	-0.25	-0.22	29.92	.000*
	Level of participation in guidance activities		0	-0.10	-0.01	5.28	.071
	Mathematics Enjoyment level		0	0.05	-0.05	24.38	.000*
	Value given to Mathematics		0	0.08	0.00	16.51	.000*
School, teacher, and classroom variables	Absenteeism		0	-0.35	0.01	30.14	.000*
	Computer use for studying at school		0	-0.24	-0.24	37.66	.000*
	In-school course (Mathematics)	Not Participated	0	0.00	0.00	0.15	.930
		Participated	0	0.03	-0.03		
	Time spent doing homework (Mathematics)		0	0.10	0.00	5.11	.078
	Attitude towards school		0	0.00	0.01	1.56	.460
	School bullying		0	-0.02	-0.02	7.85	.020*
	Mathematics Classroom climate		0	0.03	0.04	4.17	.120
	Mathematics Teacher perception		0	0.01	0.00	5.40	.067
	Mathematics Frequency of homework assignments	Never	0	0.00	0.00	30.51	.000*
		1-2 times a week	0	-0.55	-0.75		
		3 times and above a week	0	-0.09	-0.42		
Variables related to out-of-school educational opportunities	Number of books at home		0	0.27	0.14	39.15	.000*
	Computer use for studying at home		0	0.03	-0.19	11.15	.004*
	Out-of-school study (Mathematics)	Not Participated	0	0.00	0.00	18.07	.001*
		0-5 Months	0	-0.05	-0.23		
		6-10 Months	0	0.35	-0.41		
	Computer ownership status	No	0	0.00	0.00	6.75	.034*
		Yes	0	0.00	-0.34		
	Room availability	No	0	0.00	0.00	7.06	.029*
		Yes	0	0,31	-0,03		

*p<.05

As is seen in Table 5, in the Mixture OLR analysis, covariates were added individually and separately to determine the variables affecting the classes. The variables selected as the source of DIF in the model where Grade 1 was taken as the reference group are given below. Accordingly, the number of siblings (Wald=16.40; p<.05), father's education level (Wald=27.21; p<.05), and mother's education level (Wald=15.77; p<.05) variables were found to be adequate on latent class memberships. Among the variables related to personal characteristics and individual differences, the level of participation in social activities (Wald=29.92; p<.05) was significant on latent class membership. Among the variables related to school, teacher, and class, absenteeism (Wald=30,14; p<.05) was significant on latent class membership. Among the variables about out-of-school educational opportunities, the number of books at home (Wald=39.15; p<.05), participation in out-of-school math studies (Wald=18.07; p<.05) and having a personal room at home (Wald=7.06; p<.05) were significant on latent class memberships. In the last stage of the analysis, these variables, which were significant when included in the model separately and

individually and could be interpreted as a source of DIF, were included in the model together. In the last stage, the variables included in the model were significant and interpreted as explanatory variables of the moderating DIF effect of the latent class, in other words, as the source of DIF. The findings obtained are given in Table 6.

Table 6. *Covariates added together in the mixture OLR DIF analysis for item 8*

Variable Types	*Explanatory Variables*	*Grade1*	*Grade2*	*Grade3*	*Wald*	*p*
Family-related variables	Number of siblings	0	-0.13	0.08	9.34	.009*
Variables related to Personal Characteristics - Individual Differences	Level of participation in social activities	0	-0.25	-0.22	29.54	.000*
School, teacher, and classroom variables	Absenteeism	0	-0.27	0.03	18.86	.000*
Variables related to out-of-school educational opportunities	Out-of-school study (Mathematics)	0	0	0		
		0	-0.12	-0.21	10.90	.028*
		0	0.16	-0.43		
	Number of books at home	0	0.21	0.17	25.22	.000*

*p<.05

As is seen in Table 6, the variables that were included in the model together in the fourth stage of the analysis and were significant at the 0.05 level and were interpreted as the source of DIF were the number of siblings (Wald=9.34; p<0.05), level of participation in social activities (Wald=29.54; p<0.05), absenteeism (Wald=18.86; p<0.05), participation in out-of-school math studies (Wald=10.90; p<0.05) and number of books in the house (Wald=25.22; p<0.05). The findings show that the latent class's moderating effect affects the students' response patterns to item 8. Five of the 25 variables in the model established to determine the variables that affect the latent class membership play an essential role in changing this response. The five variables that play an important role in the differentiation of male and female students' responses to Item 8 are family-related variables such as the number of siblings, individual differences such as the level of participation in students' social activities, school, teacher, and classroom-related variables such as absenteeism, out-of-school educational opportunities such as participation in out-of-school math tutoring and the number of books in the house.

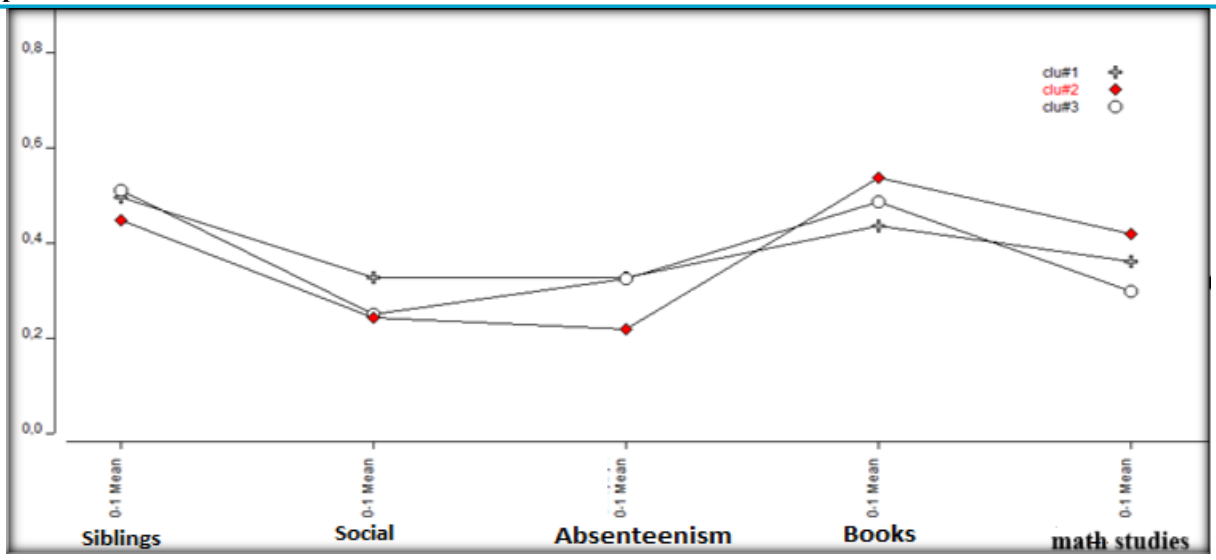

Figure 2. *Characteristics of latent class 2 in the context of variables interpreted as DIF sources*

Figure 2 shows the characteristics of the classes according to the five variables considered DIF sources. Accordingly, when the number of siblings, one of the variables considered as a source of DIF for item 8, is considered, participants in class 2 generally have fewer siblings compared to other classes. Similarly, students in class 2 participate less in social activities and are less absent. While the number of books in their homes is higher than in other classes, these students participate more in out-of-school math studies. It can be said that these variables, which are interpreted as sources of DIF, play an essential role in the differentiation of the responses of male and female students in class 2 to item 8.

Considering the analysis results for other items, the DIF effect according to gender is not significant in items 1, 2, and 6 with both OLR and Mixture OLR methods. Based on the research findings, in items 3, 4, 5, 7, and 9, which showed DIF at level A according to the OLR method, it was seen that the DIF effect was not significant based on classes after the Mixture OLR analysis, or even if it was significant on classes, the coefficients obtained for these items did not differ significantly between classes.

DISCUSSION, CONCLUSION, RECOMMENDATIONS

This study aimed to examine whether the items in the Mathematics Self-Efficacy scale of the 2016 ABIDE exam show DIF according to gender and the sources of DIF with the latent class DIF approach. The findings of the study indicated that the DIF effect in items 1, 2, and 6 of the mathematics self-efficacy scale according to gender was not significant in both the OLR and the Mixture OLR methods. On the other hand, it was concluded that in items 3, 4, 5, 7, and 9, which showed DIF at level A according to the OLR method, after the Mixture OLR analysis, the DIF effect was not significant based on classes or even if it was significant based on classes, the coefficients obtained for these items did not differ significantly between classes. In addition to this result, in item 8, it was concluded that the DIF effect obtained at level A, according to the OLR method, showed DIF at level B when the Mixture OLR method was used.

The results suggest that the number of items with DIF and the effect of DIF change when the examined group is divided into homogeneous classes. These results show the impact of homogeneity on DIF. In the literature, it is seen that Chen and Jiao (2014) stated that observed group DIF approaches assume that the observed groups (gender, race, etc.) are homogeneous, which is difficult in practice. Similarly, Samuelsen (2008) emphasized that in the observed group DIF approach, it is assumed that the groups are quite homogeneous within themselves, but the observed groups are mostly not composed of homogeneous structures. Oliveri at al. (2016) compared latent classes and observed group DIF approaches and concluded that the number of items with DIF and the size of DIF changed when the group was divided into homogeneous subclasses. Similarly, Yalçın (2017) examined the effect of latent classes formed according to students' affective characteristics on the item function differentiated by gender. These results suggest that homogeneity is an important factor in DIF analyses and that the real effect and power of DIF may not be revealed in DIF studies conducted in non-homogeneous groups.

To exemplify, item 8 was expressed as "I do not worry about failing the mathematics exam." When the responses to this item were analyzed using the OLR method, DIF was obtained at level A and in favor of males. When the moderating DIF effect of the latent class was examined, a latent structure with three classes was accepted, and a B-level DIF was obtained in favor of males in class 2. When the scores obtained from item 8 and the overall scale were analyzed, it was seen that class 2 constituted approximately 26% of the research group and that this class had high self-efficacy levels. In contrast, they scored quite low on item 8. It was concluded that five of the 25 variables added to the model as covariates to determine the characteristics of class 2, where the DIF effect was significant and that differed from the other classes where the DIF effect was not significant, were significant. According to these results, while the number of siblings, participation in social activities, and absenteeism of students in Class 2 were low, the number of books in their homes and their participation in private math tutoring outside of school were higher than the other classes.

When the findings of the latent class analysis conducted for each item are examined, it was discussed in detail whether the students' response patterns changed across classes. In this way, how the DIF effect changes according to gender in homogeneous subgroups and the sources of this change were determined. In other words, it was aimed to examine the sources of DIF. Cohen and Bolt (2005) emphasized that observed group DIF approaches can examine the existence of DIF, but they are not ideal methods for determining the sources of DIF. Similarly, Finch and Hernandez-Finch (2013) stated that a potential disadvantage of the observed group DIF approach is that the source of DIF is largely, if not entirely, based on the observed groups. In other words, a researcher examining DIF by gender assumes that DIF is based only on gender and thus does not consider other potential sources. However, the source of DIF may not be directly related to gender. Zumbo et al. (2015) stated, “the latent class DIF approach is one of the third generation DIF approaches” For example, gender should be considered a social construct in the third generation DIF approach, and “gender differences in item performance are explained by contextual or situational variables (ecological variables), such as institutionalized gender roles, class size, socio-economic status, teaching practices, and parental styles” (p. ibid). The item response pattern is examined in detail in the third generation DIF approach, and an ecological model is constructed. In other words, third generation DIF studies determine the ecology of item response by incorporating many variables such as family, in-school and out-of-school factors, social differences, teacher and classroom effects, and psychological and cognitive factors (Zumbo & Gelin, 2005).

The results for item 8 suggest that the DIF effect for this item according to gender was significant in mathematics self-efficacy scale. The DIF obtained for this item is in favor of males. In the models established with the Mixture OLR method for item 8, in the classes where DIF was significant, students generally scored high on the overall scales regardless of whether they were male or female. In contrast, they scored low on item 8. When the characteristics of the classes in which DIF was significant were examined, it was seen that the individuals in the classes in which DIF was significant had high self-efficacy scores for mathematics and low absenteeism. In addition, it was observed that the individuals in these classes had a better level of taking private lessons or courses related to this course than other classes. However, these students participate less in social activities. From this point of view, when we examined the class characteristics, it can be said that the variable that increases the level of anxiety about the exams of the participants in these classes, who are at a reasonable level according to self-efficacy and other characteristics, is their lack of socialization. The statement, "I do not worry about failing the math exam", evokes the expression of anxiety rather than self-efficacy. Nemiah (1975, as cited in Ünal-Karagüven, 1999) defined anxiety as anticipating a bad event in the future with fear. In addition, in many anxiety scales examined in the literature, it was observed that items similar to the expression of item 8 were frequently found (Bindak, 2005; Öztop, 2018).

Furthermore, the results regarding item 8 suggest that the test anxiety of the participants in the class where DIF is significant is higher than in the other classes. When gender was considered, it was concluded that female students were more anxious than male students in mathematics exams; in other words, they were more anxious. De Wit at al. (2010) stated in their meta-analysis study that physical and social

activities effectively reduce anxiety levels in individuals. When the studies on social anxiety were analyzed, it was seen that gender was an influential variable in social anxiety. In general, it was stated that social anxiety levels were higher in women than in men (Arıcıoğulları, 2001; Erözkan, 2007; Ümmet, 2007). Therefore, it can be said that social relations and activities might affect girls at a higher level. In addition to these results, although there is no consistency in the literature, some studies have reported that girls have higher math and science anxiety levels than boys (Akgün & Aydın, 2007; Kaya & Varol, 2004). Based on all these results, it can be said that female students in the class where DIF is significant have lower levels of participation in social activities compared to girls in other classes, and this situation increases their anxiety and worry levels for exams more than boys.

Based on the findings of the research, the following recommendations are presented. First, observed group DIF approaches are practical approaches for determining DIF. However, they may give biased results in cases where the examined group is not homogeneous. For this reason, examining the group's homogeneity when conducting DIF studies is recommended. Second, individuals' reactions to an item may be affected by many ecological variables such as family, teacher, school, social environment, in-school and out-of-school educational opportunities, and characteristic features such as gender. For this reason, it is suggested that third generation DIF approaches such as Mixture LR should be used in DIF studies to examine the sources of DIF that cause changes in individuals' responses in a multidimensional way. Third, according to the research results, students should emphasize social activities and studying, being interested in, valuing, and caring about school, class, and teachers. For this reason, it is recommended that students should be guided more by their families, teachers, and school administrations to participate in in-school and out-of-school social activities. Fourth, in this study, data from a 5-point Likert scale of self-efficacy were used. However, the items can be in different formats. New studies can be conducted with the Mixture LR method to determine the sources of DIF, especially in dichotomously scored tests. Last, in the study, real data belonging to the actual application of self-efficacy scales applied in ABIDE 2016 were used. In new studies, simulation studies can be conducted by considering different situations, such as multidimensionality, latent class, and overlap percentages of observed groups, to determine the power and effectiveness of the Mixed OLR method.

REFERENCES

Akgün, A., & Aydın, S. (2007). İlköğretim fen ve matematik öğretmenliği öğrencilerinin kaygı düzeylerinin bazı değişkenlere göre incelenmesi. *Elektronik Sosyal Bilimler* Dergisi, 6(20), 283-299.

Arıcıoğulları, Z., (2001). *Ortaöğretim Öğrencilerinde Sosyal Fobi ve Benlik Saygısı Arasındaki İlişki ve Bunları Etkileyen Faktörler*. (Tıpta Uzmanlık Tezi), Dicle Üniversitesi.

Bindak, R. (2005). İlköğretim öğrencileri için matematik kaygı ölçeği. *Fırat Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 17*(2), 442-448.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2014). *Bilimsel araştırma yöntemleri*. Pegem Akademi Yayıncılık.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. California: Sage.

Chen, Y. F., & Jiao, H. (2014). Exploring the utility of background and cognitive variables in explaining latent differential item functioning: An example of the PISA 2009 reading assessment. *Educational Assessment, 19(*2), 77-96. https://doi.org/10.1080/10627197.2014.903650

Cho, S. J., Suh, Y., & Lee, W. Y. (2016). An NCME Instructional Module on Latent DIF Analysis Using Mixture Item Response Models. *Educational Measurement: Issues and Practice, 35*(1), 48-61. https://doi.org/10.1111/emip.12093

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. Educational Measurement: *Issues and Practice*, 17, 31-44.

Cronbach, L. J. (1984). *Essentials of psychological testing*. Harper.

Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement, 42*(2), 133-148. https://doi.org/10.1111/j.1745-3984.2005.0000.

DeMars, C. E., & Lau, A. (2011). Differential Item Functioning Detection With Latent Classes: How Accurately Can We Detect Who Is Responding Differentially?. *Educational and Psychological Measurement, 71*(4), 597-616. https://doi.org/10.1177/001316441140422

De Ayala, R. J., Kim, S. H., Stapleton, L. M., & Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing, 2*(3), 243-276. https://doi.org/10.1080/15305058.2002.9669495.

De Wit, L. M., Fokkema, M., van Straten, A., Lamers, F., Cuijpers, P., & Penninx, B. W. (2010). Depressive and anxiety disorders and the association with obesity, physical, and social activities. *Depression and anxiety, 27*(11), 1057-1065. https://doi.org/10.1002/da.20738.

Erözkan, A. (2007). Üniversite öğrencilerinin reddedilme duyarliklari ıle sosyal kaygi düzeylerinin bazi değişkenlere göre incelenmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *1*(17), 225-240.

Finch, W. H., & Hernández Finch, M. E. (2013). Investigation of specific learning disability and testing accommodations based differential item functioning using a multilevel multidimensional mixture item response theory model. *Educational and Psychological Measurement, 73*(6), 973-993.

Frick, H., Strobl, C., & Zeileis, A. (2015). Rasch mixture models for DIF detection: A comparison of old and new score specifications. *Educational and Psychological Measurement*, *75*(2), 208-234. https://doi.org/10.1177/00131644145361.

Gomez-Benito, J., Sireci, S., Padilla, J. L., Hidalgo, M. D., & Benítez, I. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema, 30*(1), 104-109.

Hu, L. T., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118.

Kang, T., & Cohen, A. S. (2003, April). *A mixture model analysis of ethnic group DIF. Paper presented at the annual meeting of the National Council on Measurement in Education*, Chicago, IL.

Kaya, M. & Varol, K. (2004). İlahiyat fakültesi öğrencilerinin durumluk-sürekli kaygı düzeyleri ve kaygı nedenleri (Samsun örneği). *Ondokuz Mayıs Üniversitesi İlahiyat Fakültesi Dergisi, 17*(17), 31-63.

Kayri, M. (2006). *Özel yetenek sınavında (Beden Eğitimi) başarının modellenmesi ve risk faktörünün tanımlanması* [Yayımlanmamış Doktora Tezi]. Yüzüncü Yıl Üniversitesi.

Kelecioğlu, H., & Şahin, S. G. (2014). Geçmişten Günümüze Geçerlik. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 5*(2), 1-11. https://doi.org/10.21031/epod.41706

Lee Webb, M. Y., Cohen, A. S., & Schwanenflugel, P. J. (2008). Latent class analysis of differential item functioning on the Peabody Picture Vocabulary Test–III. *Educational and Psychological Measurement, 68*(2), 335-351. https://doi.org/10.1177/001316440730847

MEB (2016a). *Uluslararası öğrenci değerlendirme programı PISA 2015 ulusal raporu*. https://odsgm.meb.gov.tr/test/analizler/docs/PISA/PISA2015_Ulusal_Rapor.pdf

MEB (2016b). *TIMSS 2015 Ulusal matematik ve fen bilimleri ön raporu 4. ve 8. Sınıflar*. https://docplayer.biz.tr/130870560-Timss-uluslararasi-matematik-ve-fen-egilimleri-arastirmasi-timss-2015-ulusal-matematik-ve-fen-bilimleri-on-raporu-4-ve-8-siniflar.html

MEB (2016c). *ABİDE 2016-Akademik becerilerin izlenmesi ve değerlendirilmesi raporu*. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_11/30114819_iY-web-v6.pdf

Mislevy, R., Wilson, M., Ercikan, K., & Chudowsky, N. (2002). *Psychometric principles in student evaluation*. In D. Nevo & D. Stufflebeam (Eds.), International handbook of educational evaluation (pp. 478–520). Dordrecht, The Netherlands: Kluwer.

Oliveri, M. E., Ercikan, K., & Zumbo, B. (2013). Analysis of sources of latent class differential item functioning in international assessments. *International Journal of Testing*, *13*(3), 272-293. https://doi.org/10.1080/15305058.2012.738266.

Oliveri, M. E., Ercikan, K., Lyons-Thomas, J., & Holtzman, S. (2016). Analyzing fairness among linguistic minority populations using a latent class differential item functioning approach. Applied Measurement in Education, 29(1), 17-29.

Öztop, F. (2018). *İlkokul öğrenci velilerinin matematik kaygısının çeşitli değişkenlere göre incelenmesi. Yüksek Lisans Tezi* [Yayımlanmamış Yüksek Lisans Tezi]. Kırakkale Üniversitesi.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14* (3), 271-282. https://doi.org/10.1177/014662169001400

Samuelsen, K. M. (2005). *Examining differential item functioning from a latent class perspective*. http://gradworks.umi.com/31/75/3175148.html.

Samuelsen, K. M. (2008). *Examining differential item functioning from a latent mixture perspective*. In G. R. Hancock & K. M. Samuelsen (Eds.), Advances in latent variable mixture models (pp. 177-197). Charlotte, NC: Information Age.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Ümmet, D. (2007). *Üniversite Öğrencilerinde Sosyal Kaygının Cinsiyet Rolleri Ve Aile Ortamı Bağlamında İncelenmesi* [Yayınlanmamış Doktora Tezi]. Marmara Üniversitesi.

Ünal-Karagüven, M. H. (1999). Açık kaygı ölçeğinîn geçerlik ve güvenirliği ile ilgili bîr çalışma. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 11(11), 203-218.

Vermunt, J. K., & Magidson, J. (2005). Latent GOLD 4.0 user's guide. https://www.statisticalinnovations.com/wp-content /uploads/ LGusersguide.

Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods. Umea University*. EM No 60. http://coshima.davidrjfikis.com/EPRS9360/Articles/Wiberg_07.pdf

Yalçın, S., & Tavşancıl, E. (2015). TIMSS 2011 Fen uygulamasında cinsiyete göre farklılaşan madde fonksiyonunu açıklayan değişkenler. *Journal of Educational Sciences & Practices, 14*(27), 1-21.

Yalçın, S. (2017). Öğrencilerin Duyuşsal Özelliklerine Göre Oluşan Gizil Sınıfların Cinsiyete Göre Farklılaşan Madde Fonksiyonu'na Etkisi. *İlköğretim Online, 16*(4), 1917-1931. https://doi.org/10.17051/ilkonline.2017.364505.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF). Ottawa: National Defense Headquarters*. https://s3.amazonaws.com/academia.edu. documents/33861736/ handbook-pdf_dif. pdf

Zumbo, B. D., & Gelin, M. N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological/community moderated (or mediated) test and item bias. *Journal of Educational Research and Policy Studies, 5*(1), 1-23.

Zumbo, B. D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*(2), 223-233. https://doi.org/10.1080/15434300701375832.

Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Olvera Astivia, O. L., & Ark, T. K. (2015). A methodology for Zumbo's third generation DIF analyses and the ecology of item responding. *Language Assessment Quarterly, 12*(1), 136-151. https://doi.org/10.1080/15434303.2014.972559.

GENİŞLETİLMİŞ ÖZET

Giriş: Bu çalışmada, ABİDE sınavı kapsamında uygulanan Matematik özyeterlik ölçeğinde yer alan 9 maddelerin cinsiyete göre DMF gösterip göstermediği ve DMF kaynakları Karma OLR yöntemi ile incelenmiştir. Öğrencilerin, bir maddeye verdikleri tepkileri etkileyen birçok faktör mevcuttur. Örneğin DMF ile ilgili çalışmaların çoğunda maddede gösterilen performans farklılıkları cinsiyet ve kültür gibi karakteristik özelliklere göre karşılaştırılmış ve öğrencilerin performansını etkileyebilecek madde içeriği ve formatı gibi madde özellikleri üzerinde durulmuştur. Ancak öğrencinin madde tepki örüntüsünü etkileyebilecek farklı birçok değişken mevcutttur. Örneğin öğrencilerin sosyo-ekonomik durumu, kişisel özellikleri, okul ve sınıf özellikleri, ebeveyn tutumları ve tarzları ve farklı öğretim uygulamaları gibi faktörler de maddeye verilen tepkileri etkileyebilmektedir (Zumbo & Gelin, 2005). Bu araştırmada öğrencilerin madde tepki örüntülerini etkileyebilecek bu tür değişkenler dikkate alınarak ABİDE Matematik özeyeterlik ölçeğine verdikleri tepkiler üçüncü nesil DMF yaklaşımlarından olan Karma OLR yöntemiyle bütüncül ve detaylı bir şekilde incelenmiştir. DMF'nin varlığı ya da yokluğundan ziyade DMF kaynaklarına odaklanılması bu araştırmanın önemli yönünü ortaya koymaktadır.

Bu çalışmanın amacı, 2016 yılı ABİDE sınavı Matematik Özyeterlik ölçeğinde yer alan maddelerin cinsiyete göre DMF gösterip göstermediğinin ve DMF kaynaklarının gizil sınıf DMF yaklaşımıyla incelenmesidir. Bu temel amaç doğrultusunda aşağıdaki sorulara cevap aranmıştır;

1. 2016 ABİDE Matematik Özyeterlik ölçeğinde OLR yöntemi ile yapılan analizlerde cinsiyete göre DMF gösteren maddeler var mıdır?

2016 ABİDE Matematik Özyeterlik ölçeğini cevaplayan katılımcılar her bir madde bazında;

1.1. Ölçülen özellik açısından birden fazla gizil sınıf içermekte midir?

Eğer birden fazla gizil sınıf varsa;

1.1.1. Karma OLR yöntemine göre yapılan parametre kestirimleri (cinsiyete göre DMF) gizil sınıflarda farklılaşmakta mıdır?

1.1.2. Gizil sınıf üyelikleri üzerinde anlamlı yordayıcı olan değişken ya da değişkenler (öğretmen, okul, aile, tutum, ilgi vb. ekolojik değişkenler) var mıdır?

1.1.3. DMF'nin varlığı ve büyüklüğü kullanılan yönteme (OLR ve Karma OLR) göre değişmekte midir?

Yöntem: Bu araştırmada öğrencilerin madde tepki örüntülerini etkileyen değişkenler ele alınacağı için korelasyonel araştırma türüne göre tasarlanmıştır. ABİDE 2016 uygulamasında toplamda 35000 sekizinci sınıf öğrencisi yer almıştır. Araştırmada aynı kitapçığı alan 15000 öğrenci içinden seçkisiz olarak seçilen toplam 5000 öğrenciye ait veri kullanılmıştır. Araştırma kapsamında DMF gösteren maddelerin ve DMF kaynaklarının belirlenmesi için Zumbo (2007) ve Zumbo vd. (2015) tarafından önerilen üçüncü nesil DMF yaklaşımı yöntemlerinden Karma Lojistik Regresyon Yöntemi kullanılmıştır. Bu yöntem genel olarak dört basamaktan oluşmaktadır; ilk aşamada klasik LR analizi yapılmaktadır. İkinci aşamada madde tepkilerine ait gizil sınıf sayısı belirlenir ve her bir sınıf için aynı anda DMF analizi yapılır. Bu aşamalar 3. ve 4. aşamaya temel oluşturur. Eğer gizil sınıfın moderatör etkisi anlamlı ise bu durumda gizil sınıf üyeliklerini etkileyen değişkenler tespit edililir. Bu durumda gizil sınıflarda anlamlı olan değişkenler aynı zamanda DMF'nin kaynakları hakkında bilgi verecektir (Zumbo, 2007; Zumbo vd., 2015).

Bulgular: OLR kullanılarak yapılan analizlere göre 5 maddenin A düzeyinde DMF gösterdiği sonucuna ulaşılmıştır. Aynı maddeler Karma OLR ile incelendiğinde ise maddelerde DMF görülmemiştir. Bu sonuca ek olarak A düzeyinde DMF gösteren bir maddenin Karma OLR sonucunda DMF büyüklüğünün (B) değiştiği görülmüştür. Elde edilen sonuçlar homojenliğin DMF gösteren madde sayısı ve DMF'nin büyüklüğünü etkilediğini göstermektedir. Üç madde de her iki yönteme göre de DMF anlamlı çıkmamıştır. Araştırmada elde edilen diğer bir önemli sonuç ise gizil sınıfın moderetör etkisinin anlamlı çıktığı bir maddede (madde 8) DMF kaynağının cinsiyetten ziyade farklı ekolojik değişkenlerden kaynaklandığıdır. Gizil sınıf özellikleri incelendiğinde, özellikle DMF etkisinin anlamlı olduğu sınıftaki öğrencilerin devamsızlık durumu ve kardeş sayılarının daha düşük olduğu görülmüştür. Bununla beraber bu sınıfta yer alan öğrencilerin evde var olan kitap sayıları ve matematik için okul harici kurs alma durumlarının yüksek olduğu sonucuna ulaşılmıştır. Dikkat çekici sonuç ise bu öğrencilerin sosyal faaliyetlere daha az katılmalarıdır. Öğrencilerin maddelere verdikleri tepkileri etkileyen bu değişkenler DMF kaynağı olarak yorumlanmıştır.

Tartışma, Sonuç ve Öneriler: Elde edilen sonuçlar, homojenliğin DMF'li madde sayısı ve DMF büyüklüğü

üzerindeki etkisini göstermektedir. Nitekim madde tepki örüntülerine göre oluşturulan homojen gizil sınıflarda DMF'li madde sayısı ve DMF büyüklükleri değişmiştir. Genel olarak yapılan araştırmalarda gözlenenen grupların homojen olduğu varsayılmaktadır (cinsiyet vb.) ancak pratikte bu durum zordur (Chen ve Jiao, 2014). AİBDE Matemetik özyeterlik ölçeğinde yer alan Madde 8 için yapılan analizlerde gizil sınıfın moderatör etkisi anlamlı çıkmış ve farklı sınıflarda DMF büyüklüğü değişmiştir. Bu madde için Karma OLR yöntemi ile kurulan modellerde, DMF'nin anlamlı olduğu sınıflarda genel olarak öğrenciler kız ve erkek farketmeden ölçeklerin genelinden yüksek puan almışken madde 8'den düşük puan almışlardır. Madde 8'in içeriği incelendiğinde "Matematik sınavında başarısız olacağım diye endişelenmem" şeklinde ifade edildiği görülmektedir. Bu ifade hem olumsuz yargı içermesi hem de özyeterlikten ziyade kaygıyı çağrıştırması açısından sorunlu görülmektedir. Gizil sınıf özellikleri incelendiğinde DMF'nin anlamlı olduğu gizil sınıfdaki öğrencilerin özyeterlik genel düzeyleri yüksektir. Yine bu öğrencilerin devamsızlık durumları diğer sınıftaki öğrencilere göre düşüktür ve bu öğrenciler matematik dersiyle ilgili özel ders ya da kurs alma durumlarının diğer sınıflara göre daha iyi düzeyde olduğu görülmüştür. Dikkat çekici sonuç ise bu gizil sınıftaki öğrencilerin sosyal faaliyetlere diğer sınıflara göre daha az katılmalarıdır. Bu sonuçlar, DMF'nin anlamlı olduğu gizil sınıfta yer alan kız öğrencilerin diğer sınıftaki kız öğrencilere göre sosyal faaliyetlere daha az katıldığını göstermektedir. Benzer durum erkekler için de geçerlidir ancak kız öğrencilerin bu maddeden daha düşük puan alması başarısız olma endişelerinin daha yüksek olduğunu göstermektedir. Bu durum sosyal faaliyetlere daha az katılmanın kızları daha çok etkilediğini ve bunun sınav kaygıları ve endişeleri üzerinde etkili olabileceğini göstermektedir.

Araştırma bulgularından hareketle aşağıdaki öneriler sunulmuştur;

Alan yazında sıklıkla kullanılan Gözlenen grup DMF yöntemleri her ne kadar DMF belirlemede etkili yöntemler olsa da özellikle incelenen grupların homojen olmadığı durumlarda yanlı sonuçlara neden olabilmektedir. Buradan haraketle DMF ile ilgili çalışmalarda mutlaka grupların homojenliği dikkate alınmalıdır.

DMF çalışmalarında sıklıkla dikkate alınan cinsiyet gibi karakteristik özellikler her durumda DMF kaynağı olarak yorumlanmamalıdır. Öğrencilerin madde tepki örüntüleri okul, sosyal çevre, öğretim olanakları, aile tutumları gibi birçok farklı değişkenden etkilenebilmektedir. Bu nedenle DMF çalışmalarında Karma LR gibi üçüncü nesil DMF yaklaşımınlarının kullanılarak bireylerin tepkilerinde değişime neden olan DMF kaynaklarının çok yönlü olarak incelenmesi önerilmektedir.