

**Kocaeli Üniversitesi**

**Eğitim Dergisi**

E-ISSN: 2636-8846

2025 | Cilt 8 | Sayı 1

Sayfa: 178-204



**Kocaeli University  
Journal of Education**

E-ISSN: 2636-8846

2025 | Volume 8 | Issue 1

Page: 178-204

PISA 2015 fen okuryazarlığı maddelerinin değişen  
madde fonksiyonunun gizil sınıf yaklaşımı ile  
incelenmesi

Investigation of differential item functioning of the  
PISA 2015 science literacy items with the latent class  
approach

**Onurcan CEYHAN,**  <https://orcid.org/0000-0002-5077-6806>

Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, c.onurcan@gmail.com

**Hamide Deniz GÜLLEROĞLU,**  <https://orcid.org/0000-0001-6995-8223>

Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, denizgulleroglu@yahoo.com

Bu çalışma H. Deniz Gülleroğlu danışmanlığında Onurcan Ceyhan (2020) tarafından hazırlanan "PISA 2015 fen okuryazarlığı maddelerinin değişen madde fonksiyonunun gizil sınıf yaklaşımı ile incelenmesi" başlıklı yüksek lisans tezinden üretilmiştir.

**ARAŞTIRMA MAKALESİ**

Gönderim Tarihi	Düzeltilme Tarihi	Kabul Tarihi
7 Ağustos 2022	14 Mayıs 2024, 4 Şubat 2025	7 Mart 2025

**Önerilen Atıf**

**Recommended Citation**

Ceyhan, O., & Gülleroğlu, H. D. (2025). PISA 2015 fen okuryazarlığı maddelerinin değişen madde fonksiyonunun gizil sınıf yaklaşımı ile incelenmesi. *Kocaeli Üniversitesi Eğitim Dergisi*, 8(1), 178-204.  
<http://doi.org/10.33400/kuje.1158799>

## ÖZ

Bu çalışmanın amacı, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2015 uygulamasının fen okuryazarlığı maddeleri arasından belirlenen 11 maddenin, gizil sınıf yöntemi kullanılarak Değişen Madde Fonksiyonu (DMF) açısından incelenmesidir. Yapılan ilk incelemelerde, yanıt örüntüsü, 506 kişilik Türkiye çalışma grubu için Madde Tepki Kuramı (MTK) varsayımları olan tek boyutluluk, monotonluk ve yerel bağımsızlık varsayımlarını karşılamıştır. MTK ile yapılan tahminler sonucunda maddelerin 2 Gizil Sınıflı Karma 2PLM (Mixed IRT 2PLM with two latent classes) için uygun olduğu sonucuna varılmıştır. Analizler 2 Gizil Sınıflı Karma 2PLM temel alınarak yapılmış ve ankete katılanların yaklaşık %91'inin birinci, %9'unun ise ikinci gizli sınıfta yer aldığı belirlenmiştir. İki gizil sınıflı Karma 2 PLM temel alınarak yapılan DMF analizinde Madde 4'ün hem tek biçimli olmayan (TBO) hem de tek biçimli olan (TB) DMF gösterdiği belirlenmiştir. 5, 6, 9 ve 11 numaralı maddelerin sadece tek biçimli olan (TB) DIF gösterdiği sonucuna varılmıştır. 5, 6, 9 ve 11. maddelerin iki gizil sınıfta farklı fonksiyonlaştıkları görülmüştür. DMF içeren maddelerin gizil sınıfları açıklama gücünü incelemek amacıyla literatür taramasında DMF kaynağı olarak raporlanmış değişkenler kovaryant olarak ayrı ayrı modele dahil edilmiştir. Cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma ve okula aidiyet duygusu gibi kovaryant değişkenler modele ayrı ayrı eklendiğinde sınıflandırma doğruluğu yüzdesini arttırmıştır. Dolayısıyla bu çalışmada kovaryant değişkenin gizil değişkenin bir yordayıcısı ya da önsel bilgi kaynağı olarak görev yaptığı sonucuna ulaşılmıştır. Bu durum bireylerin fen okuryazarlığı maddelerine ilişkin tepkilerinde bu değişkenlerden farklı bir boyutun etkisi ile gizil sınıflara ayrıldıklarını göstermektedir. Yine bu çalışma gözlenen değişkene dayalı DMF kaynağının gizil değişken ile zayıf düzeyde ilişkisi olması durumunda, test sonuçlarının geçerliğini önemli ölçüde etkilemediğini göstermiştir.

**Anahtar Sözcükler:** PISA, değişen madde fonksiyonu, gizil sınıf yaklaşımı, kovaryant, karma madde tepki kuramı

## ABSTRACT

This study examines 11 items determined among the science literacy items of the PISA 2015 in terms of DIF using the latent class method. In the initial investigations, the response pattern met the assumptions of unidimensionality, monotony and local independence, which are the assumptions of the IRT for Türkiye study group consisting of 506 people. As a result of the estimations made with the IRT, it was concluded that items were suitable for the Mixed IRT 2PLM with two latent classes. The analyses were conducted based on the Mixed IRT 2PLM with two latent classes, and it was determined that approximately 91% of the respondents were in the first and 9% were in the second latent class. In the DIF analysis based on the Mixed IRT 2PLM with two latent classes, item 4 showed both non-uniform and uniform DIF. It was concluded that the items numbered 5, 6, 9, and 11 only showed uniform DIF. Item 4, 5, 6, 9, and 11 functioned differently in two latent classes. In the literature review, variables reported as DIF sources were included in the model separately as covariant variables. The percentage of classification accuracy increased among these covariant variables such as gender, ICT resources, enjoyment of science, and sense of belonging to the school. Therefore, these specified covariant variables serve as either a predictor of the latent variable or a source of a priori knowledge. Observed variables were added as covariant variables, DIF was determined as a source, and it was concluded that they were insufficient to explain latent classes. This study showed that if the DIF source based on the observed variable had a weak relationship with the latent variable, it did not significantly affect the validity of the test results.

**Keywords:** PISA, differential item functioning, latent class approach, covariant, mixture item response theory

## INTRODUCTION

Large-scale tests, widely applied today, have an essential role in countries' education planning and accountability. These practices allow countries to make crucial decisions such as creating their policies or continuing the policies implemented. Large-scale tests are prepared within the standards determined at regional, national or international levels and applied to a wide range of students.

Item bias comes forward as a prerequisite to be examined since there is a desire to measure the difference between actual knowledge and skills in large-scale assessments. Clauser and Mazor (1998) considered item bias as one of the most critical threats to validity. Item bias studies allow for not only determining systematic differences but also revealing the reasons behind these differences (Finch & French, 2019). Before revealing the reasons behind the difference, it is necessary to investigate whether the items have bias potential with the differential item functioning (DIF) analysis (McAllister, 1993).

Differential item functioning analyses are generally based on comparing groups of an observed variable such as gender, ethnicity, and country (Holand & Wainer, 1993; Thissen, Steinberg & Wainer, 1988). However, it is difficult to explain the source of DIF with these analyses. Ackerman (1992) stated that some cognitive characteristics of individuals such as test-taking speed and problem-solving strategies, can be the primary source of DIF. Therefore, the DIF detection methods based on latent traits have also emerged.

PISA (Programme for International Student Assessment), recognized as the largest international assessment of its kind, aims to support the sustainable development of participating countries by providing feedback through comparative analyses of their educational systems (Kucam & Gülleroğlu, 2023). PISA assesses the competencies of 15-year-old students in reading, mathematics, and science literacy. In 2012, PISA included 65 participating countries, including Türkiye (OECD, 2015). It is seen that the DIF analyses are generally conducted based on an observed variable in the PISA applications, and the items with DIF are determined accordingly (Akcan & Atalay-Kabasakal, 2023; Aybek, Yaşar & Kartal, 2021; Çelik & Özkan, 2020; Daşcıoğlu, & Öğretmen, T. 2024; De Ayala, Kim, Stapleton & Dayton, 2002; Kan, Sünbül & Ömür, 2013; Kucam & Gülleroğlu, 2023; Saatcıoğlu, 2022; Sadak, 2022; Samuelsen, 2005; Toprak & Yakar, 2017; Usta, 2020; Webb, 2006).

Similarly, another international assessment like PISA is TIMSS, a large-scale comparative study that evaluates student achievement in mathematics and science globally. Recent DIF studies have predominantly relied on simulated data (Cho, 2007; Cho & Cohen, 2010; de Ayala et al., 2002; Sırgancı, 2019; Uyar, 2015). Current research has identified several shortcomings, including neglecting the consideration of levelness (Toker & Green, 2021; Yalcin, 2018), conducting DIF analyses solely based on manifest groups (Aydemir, 2023; Bayram, 2024; Unal, 2023), limited use of real data (Sırgancı, 2019; Uyar, 2015).

Other large-scale assessment, like the Quantitative Ability Tests in the Academic Staff and Postgraduate Education Entrance Examination (Çepni & Kelecioğlu, 2021) or the Medical Specialization Education Entrance Examination (Bakan Kalaycıoğlu, 2022), which serves as a high-stakes national test for placing medical graduates into specialization training programs in Türkiye, demonstrate that the growing significance of nationwide assessments in DIF research in recent years.

In recent years, the focus has been on examining individuals' cognitive and psycho-social differences apart from their achievements. With the increase of such studies, DIF resources have become a matter of discussion, and it has been suggested that the source of DIF may be latent structures (Zumbo & Gelin, 2005). The main reason for this discussion is that data sets can be heterogeneous, and the trustworthy source of DIF can be ignored. Since DIF is determined using observed variables in the traditional approach, the assumption that "the participants in the subgroups are homogeneous within themselves and heterogeneous between groups" is violated

(DeAyala, Kim, Stapleton & Dayton 2002). For the statistical studies to be solid and accurate, the heterogeneous data set should be divided into homogeneous subgroups (Kayri, 2007). At the same time, it has been stated in some of these studies that the source of DIF may originate from a variable other than the grouping variable, which is considered in line with expert opinions.

According to researchers based on latent class analysis, in many studies investigating the possible sources of DIF, variables considered to be the source of DIF were not the actual source of DIF. (Ackerman 1992; Cohen & Bolt, 2005; Maij-de Meij, Kelderman & van der Flier, 2010; Samuelsen, 2005).

In addition to students' reading skills and science and mathematics literacy, data for many variables such as demographic information, socioeconomic and sociocultural backgrounds, educational backgrounds, learning strategies, attitudes, motivations, and self-efficacy, which are thought to affect student achievement, are collected in the PISA applications. One of the purposes of collecting these data is to examine the factors affecting students' achievement. The presence of DIF in a test is the most critical threat that can reduce test validity. This data from large-scale assessments is significant in determining possible DIF sources. In the DIF analyses based on an observed variable, the primary source of DIF can be missed.

Traditional methods have been favored in previous studies, but with the increasing focus on DIF studies, mixed methods are taking precedence. The limitation of traditional methods in providing a homogeneous distribution in subgroups for DIF detection has led to doubts about the accuracy of identified DIF sources. The new approach based on latent class analysis offers researchers a significant advantage in identifying genuine DIF sources.

The study specifically aims to use the latent class approach to identify potential sources of DIF in PISA 2015 science literacy items. This undertaking carries significant implications for test developers and practitioners, affording them the opportunity to factor in DIF sources during assessment design. Identifying potential DIF sources will help ensure that investments in science education are based on scientific principles and used effectively. Additionally, there is limited research on DIF analyses using the latent class approach with PISA assessment in the national literature, making this study an important resource for future researchers using the latent class method for DIF studies.

The inclusion of covariate addition in the latent class method for DIF determination will lead to a more comprehensive understanding of potential DIF sources. As previously mentioned, PISA is a widely utilized large-scale assessment tool across various educational contexts. Consequently, validating the measurement tools within the PISA framework becomes imperative. Therefore, this study aims to employ the latent class approach, incorporating covariate addition, to assess the differential functioning of science literacy test items within the PISA 2015 application. By doing so, the research aims to provide a nuanced analysis that considers both the latent class method and covariate effects in identifying DIF sources accurately.

## METHOD

In this study, items showing DIF according to the latent class approach were determined in the study group of the PISA 2015 application, and the factors affecting students' success and the state of cognitive skills explaining DIF were revealed. In this respect, the study is descriptive and suitable for a scanning model.

## Participants

In this study, the dichotomous items measuring scientific literacy in the PISA 2015 assessment were used alongside mathematics literacy and reading skills scores, as well as variables from the school and student questionnaires. Due to the matrix sampling design of the PISA assessment, in which students are not exposed to the same set of items, the items included in the DMF analysis were limited to 11 dichotomous items measuring scientific literacy. The analyses were conducted based on these selected items. This constraint was also acknowledged as a limitation

of the study. The analyses were conducted based on these selected items. This constraint was also acknowledged as a limitation of the study.

Data from 506 individuals from Türkiye who responded to a predetermined set of 11 scientific literacy items were analyzed. These individuals form the study group, rather than the overall sample of 5895 individuals from Türkiye in the PISA 2015 dataset.

The booklets numbered 31 through 48 were used in this study. Of the booklets determined, 11 items were selected to obtain the response pattern for analysis. These items are CS627Q01S (Generic ID, S627Q01), CS627Q04S (Generic ID, S627Q04), CS635Q02S (Generic ID, S635Q02), CS627Q04S (Generic ID, S635Q03), DS635Q03C (Generic ID, S635Q03), CS603Q01S (Generic ID, S603Q01), DS603Q02C (Generic ID, S603Q02), CS603Q03S (Generic ID, S603Q03), CS603Q05S (Generic ID, S603Q05), CS602Q01S (Generic ID, S602Q01), CS602Q02S (Generic ID, S602Q02) and CS602Q04S (Generic ID, S602Q04).

This study group determined from the Türkiye population of the data set of PISA 2015, which consists of 5895 people. Seventy individuals who did not respond to the specified items were excluded from the response pattern. These individuals are those who, despite participating in the PISA, did not answer one or more items. The obtained response pattern of 506 individuals was subjected to statistical analysis in accordance with the research questions. Table 1 represents the PISA 2015 overall Türkiye population and the study group in this study by gender, class, school type, and region.

**Table 1**

*Overall Türkiye Dataset Group who attended to the PISA 2015 and Distribution of the Working Group Used in the Study According to Gender, Class, And School Type*

Variable	Group	Türkiye Dataset Group		Study Group	
		f	%	f	%
Gender	Female	2938	49.84	268	52.96
	Male	2957	50.16	238	47.04
	Total	5895	100.00	506	100.00
Grade	7. Grade	16	0.27	0	0.00
	8. Grade	105	1.78	7	1.38
	9. Grade	1273	21.59	92	18.18
	10. Grade	4308	73.08	390	77.08
	11. Grade	186	3.16	16	3.16
	12. Grade	7	0.12	1	0.20
	Total	5895	100.00	506	100.00
School Type	Secondary School	121	2.05	7	1.38
	High School (Anatolian High School and High School)	3221	54.64	298	58.89
	Vocational Technical High School	2553	43.31	201	39.72
	Total	5895	100.00	506	100.00

According to Table 1, 49.84% of the Türkiye overall sample of the PISA 2015 application is female (2938 people), while 50.16% are male (2957 people). In comparison, 52.96% of the study group in this article is female (268 individuals), and 47.04% is male (228 individuals).

### Instrument and Relevant Information

The science literacy test items and school and student survey data used within the study were obtained from the OECD web page about the PISA (<http://www.oecd.org/pisa/data/2015database/>). The PISA science literacy test consists of 189 items in total. These items were



created to measure three competence areas. For this reason, 89 items were designed to measure the competence in explaining the facts scientifically, 39 items to measure the competence in designing and evaluating a scientific inquiry method, and 56 items to measure the competence in interpreting data and findings scientifically (OECD, 2017).

In the PISA 2015 application, a 35-minute student survey was delivered following the achievement tests to reveal the factors affecting students' achievement. Questions regarding the students' socioeconomic and sociocultural characteristics, educational backgrounds, learning strategies, attitudes, motivations, and self-efficacy were asked in this survey (Arici, 2019).

## Data Analysis

In this study, the binary scored items measuring science literacy of the PISA 2015 application and the school variables, and student surveys were used. While collocating the data, 11 items with binary scores that measure science literacy were determined, and the DIF analyses were made on these items. In IRT, recoding values below the mean as 0 and values above the mean as 1 serves to simplify the modeling process, especially when working with binary outcome models like the Rasch model (De Ayala, 2009). Variables that could be potential sources of DIF were transformed into categorical forms, represented as 0 and 1. To facilitate classification, the covariates used in the analysis were recoded accordingly and included in the model. While coding the variables as 1-0, firstly, the average of all data was taken, and below the mean was re-encoded as 0, and above the mean was re-encoded as 1. The variables subjected to this transformation include home educational resources (HEDRES), enjoyment of science (JOYSCIE), ICT resources (ICTRES), test anxiety (ANXTEST), sense of belonging to school (BELONG), and cultural possessions at home (CULTPOSS).

During the data analysis process, SPSS 20 package program was used for the missing data and the extreme value analyses. Mplus 7.0 was used for the factor analysis. The unidimensionality test refers to the condition where the probability of correctly answering test items depends on only one type of ability (Crocker & Algina, 1986; Embretson, 2000; Hambleton and Swaminathan, 1985; Lord & Novick, 1968; Lord, 1980). The unidimensionality of science literacy items has been examined using confirmatory factor analysis (Lord & Novick, 1968). In order to decide whether the science literacy items meet the monotonicity assumption, the item characteristic curves of each item were also examined. During the data analysis, the dataset was assessed for the assumptions of confirmatory factor analysis, univariate and multivariate normality, linearity, singularity, and multicollinearity. (Harrington, 2009). The "sirt" package in R 3.5.2 (Zitzmann, Lüdtke & Robitzsch, 2015) was used for the Q3 statistics with which local independence (Dogan & Aybek, 2021) was evaluated, and finally, Mplus 7 was used for the model estimations, the model data fit and the latent class DIF analyses.

## Examining the assumptions

The most important assumptions of the Item Response Theory models are unidimensionality, monotonicity and local independence (Hambleton & Swaminathan, 1985; Lord, 1980; Lord & Novick, 1968; Reckase, 2009). The unidimensionality of the science literacy items was examined by confirmatory factor analysis (Lord & Novick, 1968). The data set was examined in terms of the assumptions of confirmatory factor analysis which are missing values, outliers, univariate and multivariate normality, linearity, singularity and multicollinearity assumptions (Harrington, 2009). The science literacy data set was checked in terms of missing data and no missing data was encountered. In this study, the science literacy item scores were converted into the standard scores (z score) to determine univariate outliers. There was no data set whose z score was not in between  $\pm 3.3$  (Tabachnick & Fidell, 2013) range in the data. The multivariate extreme value analysis was investigated according to the Mahalanobis distances (MD) ( $p < 0.001$ ). According to Tabachnick and Fidell (2013), MD value should be compared with the value of chi square table that takes the number of independent variables as the degrees of freedom. When MD values were examined, it was seen that there was no value exceeding the critical value of chi square

31.26. Since the skewness and kurtosis values are between -2 and +2, it is assumed that the variables have a normal distribution (George & Mallery, 2010). In this study, the skewness coefficient for the science literacy items ranges from -1.346 to 1.189, and the kurtosis coefficient ranges from -1.992 to -0.188, which are within the limits of normality. It is observed that the correlation coefficients showing the relationship between the science literacy items range from 0.036 to 0.333. In line with this, it was decided that there was no singularity problem for the data set.

Multicollinearity and singularity are problems with a correlation matrix that occurs when variables are too highly correlated. With multicollinearity, the variables are very highly correlated (say, .90 and above); with singularity, the variables are redundant; one of the variables is a combination of two or more of the other variables (Tabachnick & Fidell, 2013). Since there was no correlation of .90 and above found for the items in the perception and expectation scale, it was determined that there was no problem in terms of singularity (Tabachnick & Fidell, 2013).

The CI (condition index), the VIF (variance inflation rates) and the tolerance values were calculated to examine the multiple correlations between the variables. The results of the multicollinearity analysis of the science literacy items are presented in Table 2.

**Table 2**

*Tolerance, VIF, and CI Values*

Model	Variable	Tolerance Values	VIF values	Dimension	Eigenvalues	Condition Index
1	1	.912	1.097	1	.759	1.000
	2	.837	1.194	2	.723	3.239
	3	.888	1.127	3	.666	3.375
	4	.794	1.260	4	.550	3.713
	5	.703	1.423	5	.447	4.118
	6	.801	1.249	6	.418	4.260
	7	.806	1.241	7	.379	4.475
	8	.875	1.143	8	.348	4.672
	9	.925	1.081	9	.319	4.878
	10	.876	1.142	10	.261	5.397
	11	.785	1.274	11	.198	6.193
				12	.103	8.580

When Table 2 is examined, it was seen that the VIF values of the science literacy items were less than 10, the tolerance values were different from zero, and the condition index (CI) values were below 30. Therefore, it was concluded that there was no multicollinearity problem in scales.

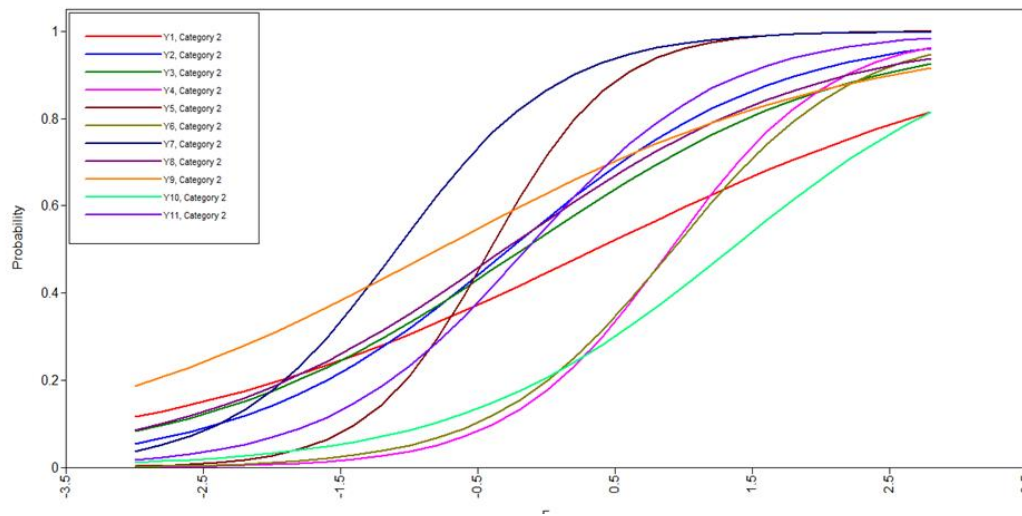
The fit of the science literacy single factor model to the data set was evaluated using the indexes of the goodness of fit. Since the chi-square statistics are affected by the sample size, the  $\chi^2/sd$  ratio, which is less affected by the sample, is a criterion that can be used instead (Waltz, Strickland & Lenz 2010). The goodness of fit indices for the single factor model of the science literacy test is presented in Table 3. When Table 3 is examined, it is seen that the single-factor structure of the science literacy test has a perfect fit.

**Table 3***Model Fit Values for the Science Literacy Test*

Scale	$\chi^2$	df	$\chi^2/df$	RMSEA	TLI	CFI	GFI
Science Literacy	61.34*	44	1.39	0.028	0.93	0.94	0.99

\* $p < 0.01$ . Note. df = degrees of freedom, RMSEA= root mean square error of approximation, NFI=Normed Fit Index, TLI= (Tucker-Lewis Index, CFI= comparative fit index, GFI=Goodness of Fit Index, PGFI = Parsimony Goodness of Fit Index.

In order to decide whether the science literacy items meet the monotonicity assumption, the item characteristic curves of each item were examined. Figure 1 represents the item characteristic curves plotted according to the two-parameter logistic model.

**Figure 1***Item Characteristic Curves for Eleven Items*

When Figure 1 is examined, it is seen that 11 items measuring the PISA science literacy are monotonically increasing. Local independence was examined with the Q3 statistics. In the 2PLM model, the correlations among the items vary between -0.201 and 0.024, according to the Q3 statistics results. DeMars (2010) states that local independence is ensured when the correlations among the items are below 0.20. Therefore, it can be said that the PISA 2015 science literacy items meet the assumption of local independence.

## Research Ethics

All the guidelines set forth in the "Ethics Directive of Higher Education Institutions Scientific Research and Publication" were followed during the entire process, from the planning of this research to its implementation, from the collecting of data to its analysis. Nothing was done that violated the guidelines outlined in the second section of the heading "Actions Contrary to Scientific Research and Publication Ethics." The participants formally agreed to participate in the study, no changes were made to the data collected, and this work was not submitted to another academic publication medium for review. Scientific, ethical, and citation norms were observed throughout the writing process.

## Research ethics committee approval information

Name of the committee that made the ethical evaluation: Ankara University Ethical Board

Date of ethical review decision: 12/07/2019

Ethics assessment document issue number: 274

Onurcan CEYHAN, Hamide Deniz GÜLLEROĞLU

Investigation of differential item functioning of the PISA 2015 science literacy items with the latent class approach



## RESULTS

In the initial phase of the research, the adequacy of different IRT models to the dataset was assessed. This assessment was guided by established criteria in the literature such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), recognized for their efficacy in evaluating model fit (Akaike, 1973; Schwarz, 1978; Spiegelhalter, Best, Carlin, & Van der Linde, 2002). The selection of the most suitable IRT or mixture (mixed) IRT model for the dataset was determined through a comparative analysis of AIC, BIC, and Log Likelihood (LL) values. Table 4 presents the model fit index values for both the IRT and Mixed IRT models. The analysis revealed that among the traditional IRT models (e.g., 2PLM, 3PLM), the two-parameter logistic model (2PLM) demonstrated superior fit based on LL, BIC, and AIC criteria. Furthermore, when exploring mixed IRT models, the mixed IRT model with two latent classes (2LC) and two parameters exhibited the best fit according to the BIC value. This evaluation process aligns with established methodologies in IRT research, ensuring that the selected model optimally captures the underlying patterns and nuances present in the dataset.

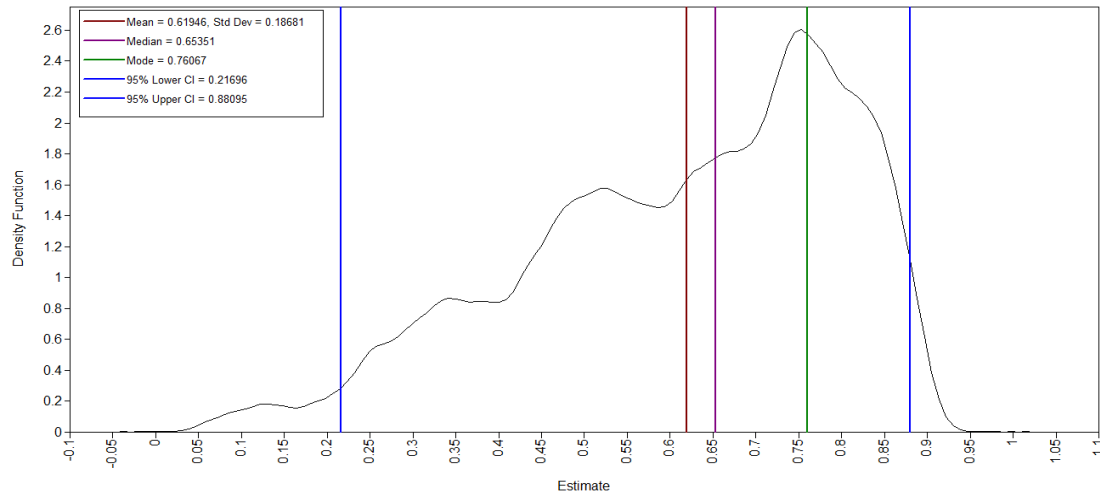
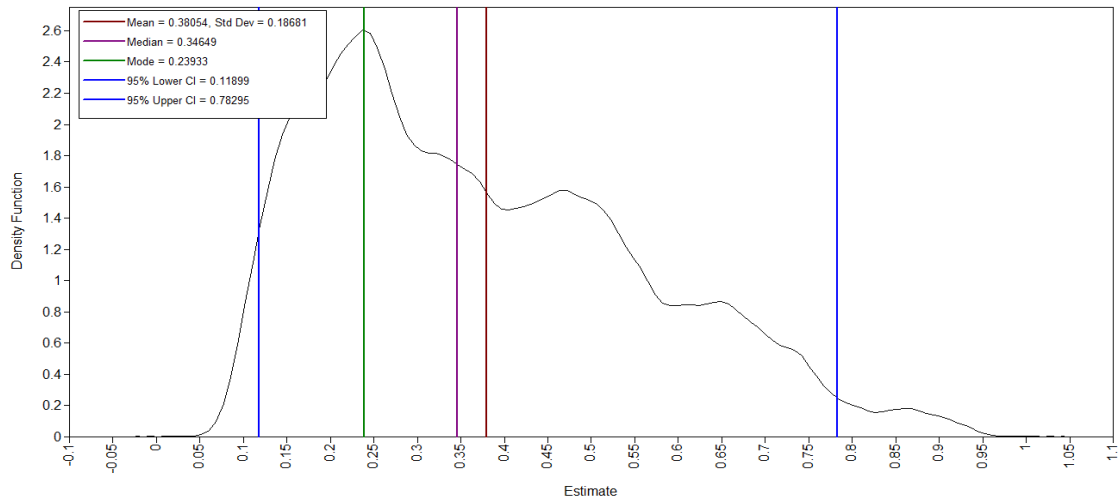
Upon making the model selection based on BIC scores, the evaluation of traditional and mixture IRT models was initiated by prioritizing the examination of Log Likelihood (LL) values among the goodness-of-fit values outlined in Table 4. The values of  $-2 \log \chi^2$  were compared to find the best fitting model among the traditional and the mixture models. When comparing the values, the difference among the LL values is compared with the chi-square table value in the relevant degree of freedom, and it is stated that the model with a small LL value fits the data better in case of a significant difference (Yalçın, 2018). In line with this, since the difference between the LL values of the 2PLM and the Mixed IRT 2 PLM with two latent classes ( $3258.054 - 3227.181 = 30.873$ ) is significantly greater than the value in the chi-square table ( $\chi^2 2 (11; 0.05) = 19.68$ ), the Mixed IRT 2PLM with two latent classes with smaller LL values was decided to be the model that best fits the data.

**Table 4**

*Model Data Fit Index Values*

Model	AIC	BIC	Df	LL
2 PLM	6560.109	6653.093	22	-3258.054
3 PLM	6572.222	6711.697	-	-3253.111
Mixed IRT (2PL) 2-Latent Class	6544.362	6734.557	45	-3227.181
Mixed IRT (2PL) 3-Latent Class	6558.155	6845.560	68	-3211.078
Mixed IRT (3PL) 2-Latent Class	6570.684	6752.425	-	-3242.342

During the second phase of the research, parameter estimation was conducted based on the model that exhibited the best fit to the dataset. Subsequently, the accuracy of classification in the estimation process was evaluated by analyzing the entropy value. It is recognized in the literature that as the entropy value approaches 1, the precision of differentiating latent classes improves significantly (Nagin, 2005). Clark and Muthen (2009) further elaborated on this, indicating that an entropy value of "0.40" indicates a low level of class distinction, "0.60" signifies a moderate level, and "0.80" signifies a high level of class distinction. As a result of the examination, the entropy value of the Mixed IRT 2PLM with two latent classes of the PISA 2015 science literacy response pattern was calculated as 0.747. This value shows that the classification accuracy is moderate. The second condition for the goodness of parameter estimation is that the Kernel density functions are unimodal. At the same time, the Kernel density plot for the classification ratio parameter indicates the model data fit (Cho, 2007; Samuelsen, 2005). The Kernel density functions regarding the classification ratio are presented in Figures 2 and 3.

**Figure 2***LC-1 Classification Ratio Kernel Density Function***Figure 3***LC-2 Classification Ratio Kernel Density Function*

When Figure 2 and Figure 3 are examined, it is seen that the Kernel functions presenting the classification rate are unimodal. In line with this, it can be said that model data fit was achieved according to the Mixed IRT 2PLM with two latent classes.

The frequency and percentage values of the response pattern were calculated in order to examine the distribution of the participants to the latent classes according to the Mixed IRT 2PLM with two latent classes, whose classification accuracy was tested. The frequency and percentage values of the distribution of this response pattern to the latent classes are presented in Table 5.

**Table 5***Distribution of Respondents to Latent Classes*

	Latent Class-1		Latent Class-2		Total	
Entropy	f	%	f	%	f	%
0.747	460	90.9	46	9.10	506	100

Table 5 shows that approximately 91% of the respondents are placed in the first latent class, and 9% are placed in the second latent class. In other words, most of the respondents are gathered in the first latent class. There is no significant difference between the ability parameters of these participants in the latent classes ( $t(-0.527) = 0.598, p > 0.05$ ). It indicates that 46 respondents with the desired feature to be measured differ from 460 respondents with the desired feature due to one or more variables or the interaction of these variables.

While determining the items with DIF, the parameter comparison method, based on the item discrimination and the item difficulty parameters estimated from each latent class, was used. Table 6 shows the results of the DIF analysis according to the item parameter estimation values of the PISA 2015 science literacy test.

**Table 6***DIF Analysis Results of Items according to Latent Classes*

Item	$\chi^2$	a1	a2	DIF Type	$\chi^2$	b1	b2	DIF Type
1	0.810	0.835	-0.043	-	1.568	0	-13.348	-
2	0.016	0.826	0.692	-	0.957	0.022	-0.789	-
3	0.010	0.437	0.537	-	0.194	-0.400	-0.052	-
4	6.688*	1.943	0.438	Non U	4.820*	0.748	2.168	U
5	0.128	1.25	1.486	-	3.925*	-0.126	-0.597	U
6	0.191	1.182	0.76	-	5.264*	0.986	0.634	U
7	1.336	0.808	1.46	-	0.856	-1.184	-0.863	-
8	0.119	0.527	0.939	-	3.133	-1.142	0.203	-
9	0.197	0.477	0.336	-	4.102*	-0.321	-1.738	U
10	0.601	0.503	0.828	-	0.358	1.437	1.136	-
11	0.920	1.436	0.573	-	135.022*	-0.205	0.312	U

$\chi^2_{(1,0.05)} = 3.84$ , Significance level: '\*': 0.05. Non U = Non Uniform, U = Uniform

According to Table 6, it is observed that the item 4 shows both non-uniform and uniform DIF. On the other hand, items 5, 6, 9, and 11 were found to show only uniform DIF. In line with this, it is seen that items 4, 5, 6, 9, and 11 function differently in two latent classes.

While trying to determine the possible sources of DIF between the latent classes determined by the Mixed 2PLM with two latent classes, the variables that were in the PISA 2015 application and were determined to be the source of DIF with traditional methods in the literature were also tested to see whether they were the traditional sources of DIF or not. Then, it was examined whether these variables explained DIF according to the latent classes. Table 7 presents the traditional DIF analysis results of the variables included in the PISA 2015 application and is discussed in this study to explain the latent DIF. The DIF analyses made according to the traditional approach were examined by The MH method (Mantel & Haenszel, 1959).

**Table 7***MH Analysis Results of PISA 2015 Variables*

Variable	Item	Alfa MH	Delta MH	Effect Size
Gender	M9	0.6303	1.0847	B
Home educational resources	-	-	-	-
Enjoyment of science	M2	1.8356	-1.4273	B
	M4	0.5493	1.4080	B
ICT Resources	M4	0.5718	1.3134	B
	M5	0.502	1.6196	C
	M6	2.1854	-1.8372	C
	M7	0.4562	1.8444	C
	M10	1.8967	-1.5043	C
	M11	3.4649	-2.9203	C
Test Anxiety	M4	1.5646	-1.0519	B
	M5	1.637	-1.1582	B
	M6	1.6418	-1.1651	B
	M11	0.6416	1.043	B
Sense of Belonging to School	M4	1.7039	-1.2523	B
Cultural possessions at home	-	-	-	-
Reading Total	M5	0.6146	1.1440	B

The variable which is considered the most in determining DIF in large-scale tests, including traditional methods, is the gender variable (Berberoğlu,1995; Öğretmen,1995; DeMars,1998; Zenisky, Hambleton & Robin, 2003; Adedoyin, 2010). Mainly that the categorical structure and cultural differences center more on gender has caused researchers to focus on gender as a source of the DIF. When the gender-based DIF results of the PISA 2015 science literacy items presented in Table 7 is examined, item 9 functions differently in gender groups. On the other hand, the difference's effect is "B" which is a moderate level in the traditional method. As a result of the latent class analysis, it was concluded that the same item had the DIF. Home educational resources (HEDRES) are another important variable that affects students' achievement (Anil,2009; Chiu and Xihua, 2008; Entorf and Tatsi, 2009; Visser, Juan & Feza, 2015). In this study, Home educational resources, which is one of the factors affecting student success in the literature, home educational resources were not found to be a DIF resource for the PISA 2015 science items that were examined.

When the enjoyment of science (JOYSCIE) variable is examined with the latent class method, it is seen that item 2 and item 4 have DIF according to this variable. In other words, the two items in the science literacy test function differently for individuals who enjoy science and those who do not. It is seen that the effect size of DIF in both items is "B", that is, the moderate level. When compared with the latent class approach, it is seen that item 4 shows DIF in both methods.

Considering ICT Resources as a variable tied to socioeconomic status, ICTRES, is another variable included in the student survey in the PISA 2015 in which expert opinions are needed to understand its role in potential DIF sources or not. When Table 7 is examined, it is seen that items 4, 5, 6, 7, 10, and 11 show potential DIF in terms of the students who have and do not have resources related to information and communication technologies. When the effect size of DIF was examined, it was found that only the 4th item had a "B", moderate level, effect size, and the other items had a large effect size. When compared with the latent class approach, it is seen that items 4, 5, 6, and 11 show potential DIF in both methods. Test anxiety is one of the variables in the PISA 2015 student survey that is likely to show DIF. This study concluded that items 4, 5, 6, and 11 in the science literacy test have potential DIF according to the test anxiety variable.

When the effect size DIF is examined, it is seen that items 4, 5, 6, and 11 show "B", that shows a moderate level of DIF. Compared with latent class analysis, all items found to contain DIF in the traditional DIF method were also found to contain DIF in the latent class method.

The students' sense of belonging to the school is considered one of the variables that predict academic achievement by researchers (Van Ryzin, Gravely & Roseth, 2009; Adelabu, 2007; Anderman, 2002; Booker, 2006; Gökdağ, 2012; Allipour & Shehni, 2016). In this study, it was found that item 5 among the science literacy items contained DIF according to the variable of sense of belonging to a school. When the effect size of DIF is examined, it is seen that this item shows "B, " which is the moderate DIF level. Item 5 showed DIF on both methods. Cultural Possessions at home (CULTPOSS) were included in the covariant model as a variable that can predict the science literacy achievement of the respondents. In this study, home educational resources are not a DIF resource for the PISA 2015 science items. The average score of reading skills is also an effective variable affecting science literacy and the respondents' ability to understand, comprehend and interpret a question. In this study, it was found that item 4 among the science literacy items contained DIF according to the average reading score variable. When the effect size of DIF is examined, it is seen that this item shows "B" which is a moderate level of DIF. When compared with latent class analysis, item 4 was found to contain DIF in both methods.

In line with the findings obtained in this study, it was revealed that some of the variables included in the PISA 2015 application predicted student achievement, while some others were the sources of DIF. When the results based on the traditional DIF analysis and the latent class approach are compared, it was seen that 1 item according to the variables is "gender", "Enjoyment of science", "sense of belonging to the school" and four items according to the variables that are " ICT resources" and "test anxiety" showed a common DIF in both methods.

In order to determine the role of the variables examined above in explaining the latent DIF, in other words, to determine the effect of variables found to be the source of DIF in the traditional approach to explaining the latent DIF, these variables were added to the mixture of 2PLM with two latent classes.

The odds ratio calculated the effectiveness of each variable in classifying latent classes. The odds ratio defines the strength of the relationship between the latent class membership and the covariant variable (Dai, 2013; Li, 2014). That the odds ratio is equal to 1 means that the covariant variable does not affect class membership, while that it is equal to 10 means that there is a strong relationship between the covariant variable and the latent class membership (Dai, 2013).

The findings regarding the latent class distinction and its accuracy that stem from adding these variables to the 2PLM mixture with two latent classes are presented respectively in Table 8.



**Table 8***Classification Accuracy of Mixture 2PLM and Distribution of Covariant in Latent Classes*

	Entropy	Odds Ratio		Latent Class-1		Latent Class-2		Total	
				f	%	f	%	f	%
Gender	0.784	1.305	Female	212	54.5	56	47.9	268	52.96
			Male	177	45.5	61	52.1	238	47.04
			Total	389	100.00	117	100.00	506	100.00
Home educational resources	0.753	1.119	Below	316	67.4	24	64.9	340	67.19
			Above	153	32.6	13	35.1	166	32.81
			Total	469	100.00	37	100.00	506	100.00
Enjoyment of science	0.789	0.903	Below	53	40.5	161	42.9	214	42.29
			Above	78	59.5	214	57.1	292	57.71
			Total	131	100.00	375	100.00	506	100.00
ICT resources	0.798	0.892	Below	406	94.0	70	94.6	476	94.07
			Above	26	6.0	4	5.4	30	5.93
			Total	432	100.00	74	100.00	506	100.00
Test anxiety	0.849	1.509	Below	184	42.5	24	32.9	208	41.11
			Above	249	57.5	49	67.1	298	58.89
			Total	433	100.00	73	100.00	506	100.00
Sense of belonging to school	0.793	1.525	Below	116	77.9	249	69.7	365	72.13
			Above	33	22.1	108	30.3	141	27.87
			Total	149	100.00	357	100.00	506	100.00
Cultural possessions at home	0.713	0.877	Below	155	45.2	79	48.5	234	46.25
			Above	188	54.8	84	51.5	272	53.75
			Total	343	100.00	163	100.00	506	100.00
Reading Skills total	0.76	1.154	Below	99	50.0	143	46.4	242	47.83
			Above	99	50.0	165	53.6	264	52.17
			Total	198	100.00	308	100.00	506	100.00

Table 8 shows the entropy value that measures classification accuracy, the odds ratio, which measures the strength of the relationship between latent classes and covariant variables, and the frequencies and percentages giving the distribution of covariant variables in latent classes. When the entropy values, which show the measure of latent class accuracy, are firstly examined, it is seen that the classification accuracy increases according to the classification accuracy of the Mixed 2PLM with two latent classes (0.747) when each variable except the cultural possessions at home variable is added to the model with a single covariant at a time. This finding is an indicator of the fact that covariant variables can be a predictor of latent classes.

When the entropy values are examined on a variable basis, the variable with the highest entropy value is the test anxiety "0.85". Latent classification accuracy increased from moderate to the high level when test anxiety was included as a covariant in the mixed 2PLM with two latent classes, which shows that test anxiety is a predictor of latent classes. In other words, including the anxiety variable into the model clarifies the latent class distinction. On the other hand, when the odds ratio, which gives the overlapping ratio of the test anxiety variable with latent classes, was examined, it was seen that the level of this overlap of test anxiety with latent classes was low, although it had a higher odds ratio compared to other variables. This finding shows that the distribution of the covariant variable to latent classes is not homogeneous. As a matter of fact, when the distribution of test anxiety variable to latent classes is examined, it is observed that individuals with low- and high-test anxiety levels are heterogeneously distributed into both classes. As a result, although the test anxiety variable increased the classification accuracy, it was

not found to be a true source of DIF among latent classes. In addition, it was observed that items 4, 5, and 6 contained DIF in both traditional and latent approaches. However, the heterogeneous distribution of the test anxiety variable among latent classes proved that a different dimension other than the individuals' anxiety levels had an effect on their reactions to the science literacy items. In addition, the fact that test anxiety increases the accuracy of the classification may indicate that test anxiety acts as a priori information source.

When Table 8 is examined, the variables with the highest entropy value following the test anxiety and variable is ICT resources, sense of belonging to the school, enjoyment of science, gender, reading skills, and home educational resources variables, respectively. Although the entropy values of the mixture 2PLM with two latent classes, in which these variables are included as covariates, have increased compared to the mixture 2PLM, the classification accuracy is still moderate (Clark; Muthen, 2009). Therefore, when each of these variables is included in the model, although the classification accuracy increases, the odds ratios of these variables show that the overlap ratio with the latent variable is low. When the distributions of these variables to the latent classes are examined, it is seen that they are heterogeneously distributed to each latent class. Therefore, these variables could not be determined as the source of DIF between latent classes.

When the cultural possessions at home variable is included in the mixture of 2PLM with two latent classes as a covariant, the classification accuracy decreased to "0.71" that is the lower limit of the moderate level. In line with this, the model data fitting the model to which this variable is added as a covariant could not be achieved. When the distribution of the variable to the latent classes is examined, it is seen that the distribution is heterogeneous.

As a result, when the variables of test anxiety, gender, ICT resources, enjoyment of science, sense of belonging to the school, and reading skills, each of which were determined as DIF sources according to the observed variable, were added as covariant variables to the mixture 2PLM with two latent classes, they did not show a homogeneous distribution in the latent classes in this study. Therefore, this finding proves the effect of a different dimension other than the variables mentioned in individuals' reactions to science literacy items, and thus the formation of latent classes. On the other hand, the increase in the classification accuracy index when the variables of gender, ICT resources, enjoyment of science, and sense of belonging to school are included in the model as a covariant variable, can be considered as an indication that covariant variables may be a predictor of the latent variable, and thus act as a prior knowledge source.

The relationship between the ability parameters estimated according to the mixture 2PLM with two latent classes and the mixed 2PLM with two latent classes in which covariant variables were added was examined. Table 9 shows the correlations between the ability parameters estimated with the mixed 2PLM with two latent classes and the model with the covariant variables.

**Table 9**

*Correlation Values between M-2PLM and M-2PLM- Covariant*

M-2PLM with Covariant Added	M-2PLM
M-2PLM-Gender	0.992**
M-2PLM- Home educational resources	0.992**
M-2PLM- Enjoyment of science	0.977**
M-2PLM- ICT Resources	0.991**
M-2PLM- Test Anxiety	0.991**
M-2PLM- Sense of belonging to school	0.965**
M-2PLM- Reading Total	0.994**

\*\* Significance level 0.01

When Table 9 is examined, it is observed that there is a positive and almost 1 correlation between the ability parameters estimated based on the mixture 2PLM with two latent classes to which the covariant is added, and the ability parameters estimated based on the mixed 2PLM with two latent classes. The presence of items exhibiting differential item functioning (DIF) in a test reduces the validity of the test scores due to systematic error interference with the scores. This validity issue signifies that one group may have an advantage or disadvantage compared to another group on the test taken. Essentially, it causes the test scores of one group to be estimated higher or lower for a reason unrelated to the measured trait. In this study, variables such as gender, ICT resources, enjoyment of science, sense of belonging to the school, and reading skills were identified as sources of DIF based on observed variables. However, including these variables in the model did not significantly affect ability estimation. Therefore, identifying a variable as exhibiting DIF through the observed variable analysis suggests that this variable may not be the true source of DIF.

## DISCUSSION and CONCLUSION

The data taken from the 11 items of the PISA 2015 science literacy test met the assumptions of uni-dimensionality, monotonicity, and local independence, which are the assumptions of the IRT for a Türkiye population study group of 506 people. The PISA 2015 science items were found to be suitable for the mixed 2PLM with two latent classes, as a result of the predictions made with the IRT and the mixture IRT models. Analyses were carried out based on the mixed 2PLM with two latent classes, and it was concluded that approximately 91% of the respondents were placed in the first latent class, and 9% were in the second latent class. The Kernel density plot regarding the classification ratio of the latent class distinction of the data set (Cho, 2007; Samuelsen 2005), which was determined as an indicator of the model data fit, according to the mixed 2PLM with two latent classes, was examined. The Kernel density plot was considered as an indicator of the model data fit, and it was found that the data set provided the model data fit according to the mixed 2PLM with two latent classes.

The PISA 2015 science literacy items of the Türkiye population study group were examined to see whether they show DIF based on the latent class method, which is the aim of the study. It was found that items 5, 6, 9, and 11 showed uniform DIF, while item 4 showed both uniform and non-uniform DIF. Therefore, it was concluded that 5 of the 11 items examined among the PISA 2015 science literacy items for Türkiye showed DIF.

In this study, "gender, ICT resources, enjoyment of science, sense of belonging to the school, and reading skill" were determined to be DIF sources according to the observed variables, and it was concluded that they were insufficient in explaining the latent classes. This situation shows that the individuals were classified into the latent classes with the effect of a different dimension other than these variables in their reactions to the science literacy items. Cohen and Bolt (2005) stated that gender was a source of DIF in their study in which they analyzed DIF according to the gender. However, as a result of latent class analysis, they stated that a cognitive feature was another source of DIF and that DIF analyses made only according to the observed group variable were insufficient in determining the main source of DIF. It was also shown in the literature that the methods based on mixture models provided better results in determining the source of DIF (Ackerman 1992, Cho, Suh&Lee, 2016; Cohen&Bolt, 2005; De Ayala et.al. 2002; Maij-de Meij, Kelderman & van der Flier, 2010; Rost, 1990; Samuelsen, 2005, Sirgancı, 2019). As a result, it was concluded that a variable determined as the source of DIF according to the observed variable may cause the main source of DIF regarding the latent classes to be ignored.

When the "gender, ICT resources, enjoyment of science, sense of belonging to the school" variables were included as separate covariant variables in the model, individuals' percentage of classification accuracy being assigned to the classes increased. Therefore, it was concluded in this study that the covariant variable served as a predictor or a priori source of information for the latent variable.

In this study, it was observed that when the variables that were DIF sources based on the observed variable were included in the mixture IRT model, the odds ratio with the latent classes was seen to be quite low. Similarly, when these variables were included in the mixture IRT model, they did not cause a significant change in the ability parameter estimation of the individuals. The goal in measurement is to estimate the true abilities of individuals with the least error. Therefore, although a variable was a source of DIF based on the observed variable, these variables were not found to be significant in determining true ability.

While this research identified potential DIF sources through individual examination by adding these sources as covariant, seeking expert opinions on these sources is also important. Future research could incorporate expert perspectives to refine the determination of potential DIF sources further, ensuring a comprehensive assessment of item and its implications on latent classes. This iterative process of evaluating potential sources one by one provides a foundation for subsequent studies to leverage expert insights in deciding DIF sources conclusively.

### **Limitations of the Study**

In this study, the items scored as 0-1-2 in the achievement test and the scores obtained from the variables measured in the student questionnaire were converted into 0-1 scoring.

A limitation of this study is that the interaction effects of the added covariates in the Mixture 2PLM approach could not be examined.

Due to the matrix sampling design of the PISA assessment, where not all students are exposed to the same set of questions, the items included in the DMF analysis were restricted to 11 dichotomous items measuring scientific literacy.

### **Acknowledgement and Support**

This study is based on the master's thesis entitled "Investigation of the changing item function of PISA 2015 science literacy items with the latent class approach", written by Onurcan Ceyhan (2020) under the supervision of H. Deniz Gülleroğlu.

### **Financial Support**

As authors, we do not declare any financial support for the process of conducting the research.

### **Statement of Contribution Rate**

The authors of the study contributed equally to all processes of the study.

### **Declaration of Conflict of Interest**

As the authors of the study, we declare that we have no declaration of interest/conflict.

### **Statement of Publication Ethics**

All the rules stated in the "Higher Education Institutions Scientific Research and Publication Ethics Directive" were followed in the entire process from the planning, implementation, data collection to the analysis of the data. None of the actions specified under the second section of the Directive, "Scientific Research and Publication Ethics Actions" have been carried out.

During the writing process of this study, scientific, ethical and citation rules were followed; no falsification was made on the collected data and this study was not sent to any other academic media for evaluation.

All participants gave their informed consent to participate in the study, and it was approved by the local ethics committee of the Ankara University Declaration.

**Research ethics committee approval information**

Name of the committee that made the ethical evaluation: Ankara University Ethical Board

Date of ethical review decision: 12/07/2019

Ethics assessment document issue number: 274

**KAYNAKÇA**

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of educational measurement*, 29(1), 67-91.
- Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: A case study from the Botswana junior certificate examination in mathematics. *Educational Research and Reviews*, 5(7), 385-399.
- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.
- Akcan, R., & Atalay-Kabasakal, K. (2023). The impact of missing data on the performances of DIF detection methods. *Journal of Measurement and Evaluation in Education and Psychology*, 14(1), 95-105. <https://doi.org/10.21031/epod.1183617>
- Allipour Birgani, S., & Shehni Yailagh, M. (2016). The causal relationship between teacher affective support with performance in English using the mediation of sense of belonging, enjoyment, hopelessness, self-efficacy, and academic effort. *Development Strategies in Medical Education*, 3(1), 49-59.
- Anderman, E. M. (2002). School effects on psychological outcomes during adolescence. *Journal of educational psychology*, 94(4), 795.
- Anil, D. (2009). Factors Effecting Science Achievement of Science Students in Programme for International Students' Achievement (PISA) in Türkiye. *Education and Science*, 34 (152), 87-100.
- Arıcı, Ö. (2019). *Investigating the factors related to Turkish students' collaborative problem solving skills with mediatin models according to PISA 2015 results* [Unpublished doctoral dissertation]. Ankara University, Ankara, Türkiye.
- Aybek, E. C., Yaşar, M., & Kartal, S. (2021). Öğretmen Yapımı Bir Testteki Maddelerin Değişen Madde Fonksiyonu Bağlamında İncelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*(52), 281-300. <https://doi.org/10.9779/pauefd.825631>
- Aydemir, F. (2023). *PISA 2018 matematik ve fen bilimleri alt testlerinde değişen madde fonksiyonunun Rasch Ağacı, Mantel-Haenszel ve Lojistik Regresyon yöntemleriyle incelenmesi* [Unpublished master's thesis]. Gazi University, Ankara.
- Bakan Kalaycıoğlu (2022). Gender-based differential item functioning analysis of the medical specialization education entrance examination. *Journal of Measurement and Evaluation in Education and Psychology*, 13(1), 1-13. <https://doi.org/10.21031/epod.998592>
- Bayram, Ö. (2024). *Bir tutum ölçeği üzerinden Mantel-Haenszel ve sıralı lojistik regresyon yöntemlerine göre değişen madde fonksiyonu incelenmesi* [Unpublished master's thesis]. Kocaeli University, Kocaeli.
- Berberoğlu, G. (1995). Differential item functioning (DIF) Analysis of computation, word problem and geometry questions across gender and SES groups. *Studies in Educational Evaluation*, 21(4), 439-56.
- Bernstein, B. B. (2000). *Pedagogy, symbolic control, and identity: theory, research, critique*. Lanham, MD: Rowman & Littlefield Publishers.
- Booker, K. C. (2006). School belonging and the African American adolescent: What do we know and where should we go?. *The High School Journal*, 89(4), 1-7.
- Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction*, 18(4), 321-336.
- Cho, S.J. (2007). *A multilevel mixture IRT model for DIF analysis* [Unpublished doctoral dissertation]. Florida State University, Florida, USA. [https://getd.libs.uga.edu/pdfs/cho\\_sun-joo\\_200712\\_phd.pdf](https://getd.libs.uga.edu/pdfs/cho_sun-joo_200712_phd.pdf). Retried Date 18.12.2019
- Cho, S.-J., & Cohen, A. S. (2010). A multilevel mixture irt model with an application to dif. *Journal of educational and behavioral statistics*, 35(3), 336-370. <https://doi.org/10.3102/1076998609353111>
- 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.



- Clark, S. L., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. <https://www.statmodel.com/download/relatinglca.pdf>
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- Çepni, Z. & Kelecioğlu, H. (2021). Detecting Differential Item Functioning Using SIBTEST, MH, LR and IRT Methods. *Journal of Measurement and Evaluation in Education and Psychology*, 12(3), 267-285. <https://doi.org/10.21031/epod.988879>
- Çelik, M., & Özer Özkan, Y. (2020). PISA 2015 matematik alt testinin cinsiyet ve bölgelere göre değişen madde fonksiyonunun incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 11(3), 283-301. <https://doi.org/10.21031/epod.715020>
- Dai, Y. (2013). A mixture Rasch model with a covariate: A simulation study via Bayesian Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, 37(5), 375-396.
- Daşcıoğlu, S., & Öğretmen, T. (2024). Detection of differential item functioning with latent class analysis: PISA 2018 mathematical literacy test. *International Journal of Assessment Tools in Education*, 11(2), 249-269. <https://doi.org/10.21449/ijate.1387041>
- 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(3-4), 243-276.
- DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education*, 11(3), 279-299.
- DeMars, C. (2010). *Item Response Theory: Understanding Statistics Measurement*. Oxford University Press.
- Doğan, C. D. & Aybek, E. C. (2021). *R-Shiny ile Psikometri ve İstatistik Uygulamaları*. Pegem Akademi.
- Entorf, H., & Tatsi, E. (2009). Migrants at school: educational inequality and social interaction in the UK and Germany, IZA DP No. 4175, Bonn, Germany.
- Finch, W. H., & French, B. F. (2019). *Educational and Psychological Measurement*. New York: Routledge.
- George, D. ve Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update* (10a ed.) Boston: Pearson.
- Gökdağ, Ö. A. (2012). *Examining Adolescents' Sense of School Belonging and Coping Strategies*. [Unpublished master's thesis]. Sakarya University, Sakarya, Türkiye.
- Hambleton, R. K. Swaminathan. (1985). *Item response theory*. Boston: Springer Science & Business Media.
- Harrington, D. (2009). *Confirmatory factor analysis*. Oxford university press, Oxford.
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical care*, 38(9), 11-28
- Holland, P. W., & Wainer, H. (1993) *Differential item functioning*. 1993 Hillsdale, Lawrence Erlbaum Associates, Inc.
- Kan, A., Sünbül, Ö., & Ömür, S. (2013). 6. - 8. Sınıf Seviye Belirleme Sınavları Alt Testlerinin Çeşitli Yöntemlere Göre Değişen Madde Fonksiyonlarının İncelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 207-222. <https://doi.org/10.17860/efd.55452>
- Kayrı, M. (2007). Two-Step Clustering Analysis in Researches: A Case Study. *Eurasian Journal of Educational Research (EJER)*, (28),36-42.
- Kucam, E., & Gülleroğlu, H. D. (2023). Examination of differential item functioning in PISA 2018 mathematics literacy test with different methods. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 128-153. <https://doi.org/10.21031/epod.1122857>
- Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ, Lawrence Erlbaum Ass.
- Lord, F.M. & Novick, M.R. (1968) *Statistical Theories of Mental Test Scores*. Addison-Wesley, Menlo Park.
- Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2010). Improvement in detection of differential item functioning using a mixture item response theory model. *Multivariate Behavioral Research*, 45(6), 975-999.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.

- M.E.B. (2016). *PISA 2015 National report*. Ankara: Ministry of National Education, General Directorate of Measurement, Evaluation and Examination Services.
- McAllister, P. (1993). Testing, DIF, and public policy. P. Holland, & H. Wainer (Editors), *Differential Item Functioning* (pp. 389-396). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nagin, D. S., (2005). *Group-based modeling of development*. Harvard University Press.
- OECD. (2015). *International large-scale assessments: Origins, growth and why countries participate in PISA*. Paris: OECD Publishing.
- OECD. (2017). *PISA 2015 Results (Volume III): Students' Well-Being*. Paris: OECD Publishing.
- Öğretmen, T. (1995). *Differential Item Functioning Analysis of the Verbal Ability Section of the First Stage of the University Entrance Examination in Türkiye*. [Unpublished master's thesis]. Middle East Technical University, Ankara, Türkiye.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- Saatçioğlu, F. M. (2022). Differential item functioning across gender with MIMIC modeling: PISA 2018 financial literacy items. *International Journal of Assessment Tools in Education*, 9(3), 631-653. <https://doi.org/10.21449/ijate.1076464>
- Samuelsen, K. M. (2005). *Examining differential item functioning from a latent class perspective* [Unpublished doctoral dissertation]. University of Maryland, Maryland, USA.
- Sadak, M. (2022). TIMSS 2015 ve 2019 Matematik Sorularının Türkiye'de Cinsiyete Göre Madde Yanlılıklarının İncelenmesi: SIBTEST Prosedürü ile Değişen Madde Fonksiyonu Analizi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 23(Özel Sayı), 211-258. <https://doi.org/10.29299/kefad.961858>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sırgancı, G. (2019). *The Effect of Covariant Variable on Determination of Differential Item Functioning Using Mixture Rasch Model* [Unpublished doctoral dissertation/master's thesis]. Ankara University, Ankara, Türkiye.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.
- Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*, 6th International edition (cover) edn. Boston, [Mass.].
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. *Test Validity*. Hillsdale, NJ: Erlbaum, pg. 147-169
- Toker, T. & Green, K. (2021). A comparison of latent class analysis and the mixture rasch model using 8th grade mathematics data in the fourth international mathematics and science study (timss-2011), *International Journal of Assessment Tools in Education* 8(4), 959-974
- Toprak, E. ve Yakar, L. (2017). SBS 2011 Türkçe Alt Testindeki Maddelerin Değişen Madde Fonksiyonu Açısından Farklı Yöntemlerle İncelenmesi, *International Journal Of Eurasia Social Sciences*, Vol: 8, Issue: 26, pp. (220-231).
- Unal, F. (2023). *Farklı oranlardaki kayıp verilere farklı atama yöntemleriyle veri atamanın madde tepki kuramına dayalı yöntemlerle değişen madde fonksiyonuna etkisinin incelenmesi* [Unpublished master's thesis]. Akdeniz University, Antalya.
- Usta, G. (2020). Sınav Kaygı Ölçeği Maddelerinin Çeşitli Yöntemlere Göre Değişen Madde Fonksiyonlarının İncelenmesi. *Cumhuriyet Uluslararası Eğitim Dergisi*, 9(4), 1225-1242. <https://doi.org/10.30703/cije.703337>
- Uyar, Ş. (2015). *Gözlenen gruplara ve örtük sınıflara göre belirlenen değişen madde fonksiyonunun karşılaştırılması* [Unpublished doctoral dissertation]. Hacettepe University, Ankara.
- Van Ryzin, M. J., Gravely, A. A., & Roseth, C. J. (2009). Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being. *Journal of Youth and Adolescence*, 38(1), 1-12.
- Visser, M., Juan, A., & Feza, N. (2015). Home and school resources as predictors of mathematics performance in South Africa. *South African Journal of Education*, 35(1), 28-36.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). Measurement reliability. *Measurement in Nursing and Health Research*, 82(12)145-162.

- Webb, N. L. (2006). *Identifying content for student achievement tests. Handbook of test development*, 155-180.
- Yalçın, S. (2018). Data fit comparison of mixture item response theory models and traditional models. *International Journal of Assessment Tools in Education*, 5(2), 301-313.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items. Chicago: Scientific Software.
- Zitzmann, S., Lüdtke, O., & Robitzsch, A. (2015). A Bayesian approach to more stable estimates of group-level effects in contextual studies. *Multivariate Behavioral Research*, 50(6), 688-705.
- 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 & Policy Studies*, 5(1), 1-23.

## GENİŞLETİLMİŞ ÖZ

### Giriş

Günümüzde yaygın bir şekilde uygulanmakta olan geniş ölçekli testler, ülkelerin eğitim planlamalarında ve hesap verebilirliklerinde önemli bir role sahiptir. Bu uygulamalar, ülkelerin kendi politikalarını oluşturmalarına ya da uygulanan politikaları devam ettirme gibi önemli kararların verilebilmesine olanak sağlamaktadır. Geniş ölçekli testler bölgesel, ulusal veya uluslararası düzeyde belirlenmiş standartlar dâhilinde hazırlanarak, geniş bir öğrenci kitlesi üzerinde uygulanan testlerdir.

Geniş ölçekli sınavlarda gerçek bilgi ve beceri farkı ölçülmek istendiği için, madde yanlılığı incelenmesi gereken bir ön şart olarak karşımıza çıkmaktadır. Clauser ve Mazor, madde yanlılığını geçerliğe yönelik en önemli tehditlerden biri olarak belirtmiştir. Madde yanlılığı çalışmaları sadece sistematik farklılıkları belirlemenin dışında, bu farklılığın nedenlerinin de ortaya çıkarılmasına olanak sağlamaktadır. Farklılığın nedenlerini ortaya koymadan önce maddelerin yanlılık potansiyeli içerip içermediğinin Değişen Madde Fonksiyonu (DMF) analizi ile araştırılması gerekmektedir.

Değişen madde fonksiyonu analizleri genellikle cinsiyet, etnik köken, ülke gibi gözlenen bir değişkenin gruplarının karşılaştırılmasına dayanmaktadır. Ancak bu analizler ile DMF'nin kaynağını açıklamak güçtür. Ackerman, bireylerin test alma hızı, problem çözme stratejisi gibi bir takım bilişsel özelliklerinin DMF'nin asıl kaynağı olabileceğini dile getirmiştir. Bu nedenle, gizil özelliklere dayalı DMF belirleme yöntemleri de gündeme gelmiştir.

Son yıllarda başarı dışında bireylerin bilişsel ve psiko-sosyal farklılıklarının da incelenmesine ağırlık verilmiştir. Bu tip çalışmaların artmasıyla beraber DMF kaynakları tartışma konusu haline gelmiş hatta DMF'lerin kaynağının gizil yapılar olabileceği öne sürülmüştür. Bu tartışmanın asıl sebebi ise veri setlerinin heterojen olabileceği ve DMF'nin gerçek kaynağının göz ardı edilebileceğidir. Geleneksel yaklaşımda DMF, gözlenen değişkenler kullanılarak belirlendiğinden, alt gruplarda yer alan katılımcıların "kendi içinde homojen, gruplar arasında ise heterojen olduğu" varsayımı ihlal edilmiş olur. Yapılan istatistikî çalışmaların güçlü ve hatasız olması için heterojen olan veri setinin homojen alt gruplara bölünmesi gerekir. Homojen alt grupların oluşturulabilmesi için ise gizil sınıflara göre DMF belirleme yöntemi kullanılabilir.

PISA uygulamalarında DMF analizlerinin genellikle gözlenen bir değişken temel alınarak yapıldığı ve buna göre DMF içeren maddelerin belirlendiği görülmektedir. Aynı zamanda bu çalışmaların bazılarında DMF'nin kaynağının, uzman görüşleri doğrultusunda ele alınan grupta değişkeninin dışında başka bir değişkenden kaynaklı olabileceği de belirtilmiştir.

DMF'nin olası kaynaklarının araştırıldığı birçok çalışmada, gözlenen değişkenlerle DMF belirleme çalışmaları sonucunda DMF kaynağı olduğu düşünülen değişkenlerin, gizil sınıf analizine dayalı yöntemlerde aslında DMF'nin bir kaynağı olmadıkları belirlenmiştir. PISA uygulamalarında öğrencilerin okuma becerileri ile fen ve matematik okuryazarlıklarının yanı sıra öğrenci başarısını etkileyebileceği düşünülen demografik bilgiler, sosyoekonomik ve sosyokültürel altyapıları, eğitim geçmişleri, öğrenme stratejileri, tutumları, güdüler, öz yeterlikleri gibi birçok değişken için veri toplanmaktadır. Bu verileri toplamaktaki amaçlardan biri de öğrencilerin başarısını etkileyen faktörleri incelemektir. Bir testte DMF varlığı test geçerliğini düşürebilecek en önemli tehdittir. Geniş ölçekli sınavlardan elde edilen bu tip veriler olası DMF kaynaklarının belirlenmesi açısından önemlidir. Gözlenen bir değişkene dayalı yapılan DMF analizlerinde DMF'nin asıl kaynağı gözden kaçırılabilir. Bu nedenle gizil sınıf yöntemine dayalı DMF belirleme, olası DMF kaynaklarının daha hatasız belirlenmesini sağlayacaktır. Yukarıda da belirtildiği gibi PISA, eğitimde çok geniş yelpazede kullanılan geniş ölçekli bir uygulamadır. Dolayısıyla PISA'da yer alan ölçme araçlarının geçerliğinin ortaya konması önemlidir. Bu araştırmanın amacı, PISA 2015 uygulamasında yer alan fen okuryazarlığı test maddelerinin gizil sınıf yaklaşımı ile değişen madde fonksiyonunun belirlenmesidir.

## Yöntem

### *Çalışma tasarımı ve katılımcılar*

Yapılan veri incelemesinde 5895 kişilik evrenden en uygun yanıt örüntüsüne sahip kitapçıklar belirlenmiştir. Bunlar 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47 ve 48 no'lu kitapçıklardır. Belirlenen kitapçıklardan uygulama yapılacak yanıt örüntüsünü elde etmek için 11 madde seçilmiştir. Bu sorular CS627Q01S (Generic ID, S627Q01), CS627Q04S (Generic ID, S627Q04), CS635Q02S (Generic ID, S635Q02), CS627Q04S (Generic ID, S635Q03), DS635Q03C (Generic ID, S635Q03), CS603Q01S (Generic ID, S603Q01), DS603Q02C (Generic ID, S603Q02), CS603Q03S (Generic ID, S603Q03), CS603Q05S (Generic ID, S603Q05), CS602Q01S (Generic ID, S602Q01), CS602Q02S (Generic ID, S602Q02) ve CS602Q04S (Generic ID, S602Q04) dır. Belirlenen sorularda cevap vermeyen 70 kişi yanıt örüntüsünden çıkarılmıştır. Bu kişiler testi almasına girmesine rağmen 1 ya da daha fazla soruyu cevaplamayan kişilerdir. Elde edilen 506 kişilik yanıt örüntüsü araştırma sorularına uygun bir şekilde istatistiki analizlere tabi tutulmuştur. PISA 2015 uygulaması Türkiye örnekleminin %49.84'ü kadın iken (2938 kişi), %50.16'sı erkektir (2957 kişi). Bu çalışmada kullanılan çalışma grubu ise %52.96'sı kadın, (268 kişi), %47.04'ü erkektir (228 kişi).

### *Verilerin elde edilmesi ve toplanması*

Çalışmada kullanılan fen okuryazarlığı test sonuçları, okul ve öğrenci anketleri verileri OECD'nin PISA ile ilgili web sayfasından edinilmiştir.

PISA fen okuryazarlığı testi, toplamda 189 maddeden oluşmaktadır. Bu maddeler üç yeterlilik alanını ölçmek amacıyla oluşturulmuştur. Bu nedenle, 89 madde olguları bilimsel olarak açıklama yeterliliği, 39 madde bilimsel sorgulama yöntemi tasarlama ve değerlendirme yeterliliği ve 56 madde de veri ve bulguları bilimsel olarak yorumlama yeterliliğini ölçmek amacı ile tasarlanmıştır.

PISA 2015 uygulamasında, öğrencilerin başarılarını etkileyen faktörleri ortaya koymak amacıyla, başarı testlerini takiben 35 dakikalık öğrenci anketi uygulanmıştır. Bu ankette öğrencilere ilişkin sosyoekonomik ve sosyokültürel özellikler, eğitim geçmişleri, öğrenme stratejileri, tutumları, güdüler, öz yeterlikleri ile ilgili sorular sorulmuştur.

### *Verilerin çözümlenmesi*

Bu çalışmada, PISA 2015 uygulamasının fen okuryazarlığını ölçen iki kategorili maddeler, okul ve öğrenci anketinde yer alan değişkenler kullanılmıştır. Veriler hazırlanırken fen okuryazarlığını ölçen iki kategorili 11 madde belirlenerek DMF analizleri bu maddeler üzerinden yapılmıştır. DMF kaynağı olabilecek değişkenler 1-0 biçiminde yeniden kodlanarak, kategorik hale getirilmiştir. 1-0 biçiminde kodlamada tüm verinin ortalaması alınmış, ortalamanın altı 0, üstü ise 1 olarak yeniden kodlanmıştır.

### *Varsayımların incelenmesi*

Madde Tepki Kuramı modellerinin en önemli varsayımları tek boyutluluk, monotonluk ve yerel bağımsızlıktır. Fen okuryazarlığı maddelerinin tek boyutluluğu, doğrulayıcı faktör analizi ile incelenmiştir. Veri seti, doğrulayıcı faktör analizi sayıtlarından olan kayıp veri, tek ve çok yönlü uç değer, tek ve çok değişkenli normallik, doğrusallık, tekli ve çoklu bağlantılılık sayıtları açısından incelenmiştir. Fen okuryazarlığı veri seti kayıp veri açısından incelenmiş ve kayıp veriye rastlanmamıştır. Bu çalışmada, tek değişkenli uç değerleri belirlemek için fen okuryazarlığı madde puanları standart puana (z puanı) dönüştürülmüştür. Veri setinde z puanı  $\pm 3.3$  aralığı dışında hiçbir veri bulunmamaktadır. Çok değişkenli uç değer analizi Mahalanobis uzaklıklarına (MU) göre ( $p < 0.001$ ) araştırılmıştır. Tabachnick ve Fidell (2013)'e göre MU değerinin, bağımsız değişken sayısını serbestlik derecesi alan  $\chi^2$  tablo değeriyle karşılaştırılması gerekir. MU değerleri incelendiğinde  $\chi^2$  31.26 kritik değerini aşan değer bulunmamaktadır. Çarpıklık ve basıklık değerleri -2 ile +2 arasında olduğundan, değişkenlerin normal dağılıma



sahip olduğu varsayılmaktadır (George ve Mallery, 2010). Fen okuryazarlığı maddelerine ilişkin çarpıklık katsayısı -1,346 ile 1,189; basıklık katsayısı ise -1,992 ile -0,188 aralığında değişmekte olup, normallik sınırları içindedir. Fen okuryazarlığı maddeleri arasındaki ilişkiyi gösteren korelasyon katsayılarının 0.036 ile 0.333 arasında değiştiği görülmektedir. Bu doğrultuda veri grubu için tekliklik sorununun olmadığına karar verilmiştir. Değişkenler arasındaki çoklu bağlantılılığın incelenmesi için CI (koşul indeksi), VIF (varyans şişirme oranları) ve tolerans değerleri incelenmiştir.

Fen okuryazarlığı maddelerinin VIF değerlerinin 10'dan küçük, tolerans değerlerinin sıfırdan farklı ve koşul indeksleri (CI) değerlerinin de 30'un altında olduğu görülmektedir. Dolayısıyla ölçeklerde çoklu bağlantılılık sorununun olmadığı sonucuna ulaşılmıştır. Doğrusallık varsayımı, ölçekteki maddeler arası korelasyonlar ile incelenmiştir. Algı ve beklenti ölçek maddelerinde .90 ve üstü korelasyon saptanmadığından doğrusallık açısından herhangi bir problemin olmadığı tespit edilmiştir.

Fen okuryazarlığı tek faktörlü modelinin veri setine uyumu, uyum iyiliği indeksleri aracılığıyla değerlendirilmiştir. Ki-kare istatistiği örneklem büyüklüğünden etkilendiği için örneklemden daha az etkilenen  $\chi^2/sd$  oranı bunun yerine kullanılabilecek bir ölçüttür. Tablo 3'de fen okuryazarlığı testinin tek faktörlü model uyum iyiliği değerleri verilmiştir.

Tablo 3 incelendiğinde, fen okuryazarlığı testinin tek faktörlü yapısının mükemmel bir uyuma sahip olduğu görülmektedir. Fen okuryazarlığı maddelerinin monotonluk varsayımını karşılayıp karşılamadığına karar vermek için, her bir maddenin madde karakteristik eğrileri incelenmiştir. PISA fen okuryazarlığını ölçen 11 maddenin monoton artan olduğu görülmektedir. Bu çalışmada yerel bağımsızlık Q3 istatistiği ile incelenmiştir. 2PLM modelde, Q3 istatistik sonuçlarına göre maddeler arasındaki korelasyonlar -0.201 ile 0.024 arasında değişmektedir. DeMars (2010), maddeler arasındaki korelasyonların 0.20'nin altında olduğu durumda yerel bağımsızlığın sağlandığını belirtmiştir. Buna göre, PISA 2015 fen okuryazarlığı maddelerinin yerel bağımsızlık varsayımını sağladığı söylenebilir.

### **Verilerin analizi**

Veri analizi sürecinde, eksik veri ve uç değer analizleri için SPSS 20 paket programı kullanılmıştır. Faktör analizi için ise Mplus 7.0 kullanılmıştır. Fen okuryazarlığı maddelerinin tek boyutluluğu, doğrulayıcı faktör analizi ile incelenmiştir (Lord ve Novick, 1968). Fen okuryazarlığı maddelerinin monotonluk varsayımını karşılayıp karşılamadığını belirlemek için her bir maddenin madde karakteristik eğrileri de incelenmiştir. Veri analizi sürecinde, doğrulayıcı faktör analizi varsayımları, tek değişkenli ve çok değişkenli normallik, doğrusal ilişki, tekliklik ve çoklu bağlantı (multicollinearity) açısından veri seti değerlendirilmiştir (Harrington, 2009). Yerel bağımsızlığın değerlendirilmesi için Q3 istatistikleri R 3.5.2'deki "sirt" paketi (Zitzmann, Lüdtke ve Robitzsch, 2015) kullanılarak hesaplanmış, son olarak model tahminleri, model-veri uyumu ve gizil sınıf DIF analizleri için Mplus 7 kullanılmıştır.

### **Bulgular**

Araştırmanın ilk aşamasında, veri setinin hangi MTK modeline uyum sağladığı incelenmiştir. Uyum iyiliği değerlerinden öncelikle AIC değerleri incelenerek, veri setinin geleneksel ve karma MTK modellerinden hangisine uyum sağladığı incelenmiştir. En küçük AIC değerine sahip olan model, veri setine en iyi uyum gösteren modeldir. Buna göre; geleneksel modelde 2PLM'nin, karma modelde ise 2 Gizil Sınıflı Karma 2 PLM'nin veri setine en iyi uyumu gösterdiği görülmüştür. Geleneksel ve karma modeller arasında en iyi uyum gösteren modeli bulmak amacıyla  $-2 \log \chi^2$  değerleri karşılaştırılmıştır. Değerler karşılaştırılırken LL değerleri arasındaki fark ilgili serbestlik derecesindeki kare tablo değeri ile karşılaştırılır ve anlamlı farkın olması durumunda küçük LL değerine sahip modelin veriye daha iyi uyum sağladığı belirtilir. Buna göre; 2PLM'nin ve 2 Gizil Sınıflı Karma 2 PLM'nin LL değerleri arasındaki fark  $(3258.054 - 3227.181 = 30,873)$  ki kare tablosundaki değerden  $(\chi^2 (11; 0.05) = 19.68)$  anlamlı derecede

büyük olduğundan, daha küçük LL değerine sahip 2 Gizil Sınıflı karma 2PLM'nin veriye en iyi uyum sağlayan model olduğuna karar verilmiştir.

Araştırmanın ikinci aşamasında veri setine en iyi uyum sağlayan modele göre parametre kestirimi yapılmıştır. Kestirimin sınıflama doğruluğu entropi değeri ile incelenmiştir. Entropi değeri bire yaklaştıkça gizil sınıf ayrımının doğruluğu artar. Clark ve Muthen entropi değerinin "0.40" düşük, "0.60" orta ve "0.80" yüksek sınıf ayrımını temsil ettiğini ortaya koymuştur. Yapılan inceleme sonucunda, PISA 2015 fen okuryazarlığı yanıt örüntüsünün 2 Gizil Sınıflı Karma 2PLM'nin entropi değeri 0.747 olarak hesaplanmıştır. Bu değer sınıflama doğruluğunun orta düzeyde olduğunu göstermektedir. Parametre kestirim iyiliğinin ikinci koşulu ise Kernel yoğunluk fonksiyonlarının tek modlu olmasıdır. Aynı zamanda, sınıflama oranı parametresine ilişkin Kernel yoğunluk grafiği model veri uyumunun bir göstergesidir.

Sınıflama oranını veren Kernel fonksiyonlarının tek modlu olduğu görülmektedir. Buna göre, 2 Gizil Sınıflı Karma 2 PLM 'e göre model veri uyumunun sağlandığı söylenebilir.

Yanıtlayıcıların yaklaşık %91'nin birinci gizil sınıfta, %9'unun da ikinci gizil sınıfta yer aldığı görülmektedir. Diğer bir ifade ile yanıtlayıcıların büyük çoğunluğu birinci gizil sınıfa yığılmıştır. Gizil sınıflardaki bu katılımcıların yetenek parametreleri arasında manidar bir fark bulunmamaktadır ( $t(-0.527) = 0.598, p > 0.05$ ). Bu bulgu, ölçülmek istenen özelliğe sahip 46 yanıtlayıcının, yine ölçülmek istenen özelliğe sahip 460 yanıtlayıcıdan bir ya da birden fazla değişkenden ya da bu değişkenlerin etkileşiminden dolayı farklılaştığını işaret etmektedir.

DMF gösteren maddeler belirlenirken her bir gizil sınıftan kestirilen madde ayırt edicilik ve madde güçlük parametrelerinin karşılaştırılmasına dayalı olan parametre karşılaştırma yöntemi kullanılmıştır. Madde 4'ün hem tek biçimli olmayan (TBO) hem de tek biçimli (TB) DMF gösterdiği görülmüştür. Madde 5, 6, 9 ve 11'in ise sadece TB DMF gösterdiği bulunmuştur. Buna göre madde 4, 5, 6, 9 ve 11'in iki gizil sınıfta farklı fonksiyonlaşmakta olduğu görülmektedir.

2 Gizil Sınıflı Karma 2 PLM ile belirlenen gizil sınıflar arasında DMF gösteren maddelerin olası DMF kaynaklarının belirlenmesine çalışılırken, öncelikle PISA 2015 uygulamasında yer alan ve alan yazınında geleneksel yöntemler ile DMF kaynağı oldukları belirlenen değişkenlerin bu çalışmada da geleneksel DMF kaynağı olup olmadıkları test edilmiştir. Ardından, bu değişkenlerin gizil sınıflara göre DMF'yi açıklayıp açıklamadıkları incelenmiştir.

Geniş ölçekli testlerde DMF belirlemede geleneksel yöntemler de dâhil olmak üzere en çok dikkate alınan değişken cinsiyet değişkeni olmuştur. Özellikle kategorik yapısı ve kültürel farklılıkların cinsiyet üzerinde yoğunlaşması, araştırmacıların DMF kaynağı olarak cinsiyete yönelmelerine neden olmuştur. PISA 2015 fen okuryazarlığı maddelerinin cinsiyete göre DMF sonuçları incelendiğinde, madde 9'un cinsiyet gruplarında farklı fonksiyonlaştığı görülmektedir. Geleneksel yöntemde ise farklılığın etkisinin "B", yani orta büyüklükte olduğu görülmektedir. Gizil sınıf analizi sonucunda da aynı maddenin DMF içerdiği sonucuna ulaşılmıştır. Evdeki eğitim kaynakları (HEDRES) öğrencilerin başarılarını etkileyen önemli değişkenlerden bir diğeridir. Bu çalışmada, alan yazınındaki çalışmaların aksine evdeki eğitim kaynaklarının incelenen PISA 2015 fen maddeleri için bir DMF kaynağı olmadığı sonucuna ulaşılmıştır.

Bu çalışmada elde edilen bulgular doğrultusunda, PISA 2015 uygulamasında da yer alan değişkenlerin bazılarının öğrenci başarısını yordadığı, bazılarının ise DMF kaynağı oldukları ortaya konmuştur. Geleneksel DMF analizi ile gizil yaklaşıma dayalı DMF analiz sonuçları kıyaslandığında; her iki yöntemde de "cinsiyet", "fen öğrenmekten zevk alma" ve "okula ait hissetme" değişkenlerine göre bir, "bilgi ve iletişim teknolojileri (BİT) kaynakları" ve "test kaygısı" değişkenlerine göre dört madde ortak olarak DMF göstermiştir.

Gizil DMF'yi açıklamada yukarıda incelenen değişkenlerin rolünün belirlenmesi, diğer bir ifade ile geleneksel yaklaşım ile DMF kaynağı olduğu tespit edilen değişkenlerin gizil DMF'yi açıklamadaki etkisinin belirlenmesi için bu değişkenler kovaryant olarak 2 Gizil Sınıflı Karma 2 PLM'ye eklenmiştir.

Her bir değişkenin gizil sınıflara ayırmada ne derece etkili olduğu odds oranı ile hesaplanmıştır. Odds oranı gizil sınıf üyeliği ile kovaryant değişken arasındaki ilişkinin gücünü tanımlar. Odds oranının 1'e eşit olması kovaryant değişkenin sınıf üyeliği üzerinde etkisi olmadığı; 10'a eşit olması ise kovaryant değişkenle gizil sınıf üyeliği arasında güçlü bir ilişki olduğu anlamına gelmektedir.

Sonuç olarak bu çalışmada; her biri gözlenen değişkene göre DMF kaynağı olarak belirlenen kaygı, cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma, okula ait hissetme ve okuma becerisi değişkenleri 2 Gizil Sınıflı Karma 2 PLM'ye kovaryant değişken olarak eklendiklerinde gizil sınıflarda homojen bir dağılım göstermemişlerdir. Dolayısıyla bu durum fen okuryazarlığı maddelerine bireylerin tepkilerinde bahsi geçen değişkenlerden farklı bir boyutun etkisi, dolayısıyla gizil sınıfların oluştuğuna bir kanıt oluşturmaktadır. Öte yandan cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma, okula ait hissetme değişkenlerinin modele kovaryant değişken olarak dâhil edildiğinde sınıflama doğruluğu indeksinin yükselmesi ise kovaryant değişkenlerin gizil değişkenin bir yordayıcısı olabileceği, dolayısıyla önsel bilgi kaynağı olarak görev yaptığının bir göstergesi olarak yorumlanabilir.

Kovaryantın eklendiği her bir 2 Gizil Sınıflı Karma 2 PLM 'ye göre kestirilen yetenek parametreleri ile 2 Gizil Sınıflı Karma 2 PLM'den kestirilen yetenek parametreleri arasındaki ilişki pozitif ve neredeyse 1'e yakındır. Bir testte değişen madde fonksiyonu içeren maddelerin olması test puanlarına sistematik hata karışmasından dolayı test puanlarının geçerliğini düşürür. Bu geçerlik sorunu alınan test için bir grubun diğer gruba göre avantajlı ya da dezavantajlı olduğunun göstergesidir. Diğer bir ifade ile ölçülen özelliğin dışında bir nedenden dolayı bir grubun test puanlarının daha yüksek veya düşük kestirilmesine sebep olmaktadır. Bu çalışmada cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma, okula ait hissetme ve okuma becerisi gözlenen değişkene göre DMF kaynağı olarak belirlenmiştir. Ancak bu değişkenlerin modele dâhil edilmesinin yetenek kestirimine anlamlı bir etkisi olmamıştır. Dolayısıyla gözlenen değişkene göre yapılan DMF analizleri sonucunda değişkenin DMF gösterdiğinin belirlenmesi o değişkenin gerçek DMF kaynağı olmayabileceğinin bir göstergesidir.

## Tartışma ve Sonuç

PISA 2015 fen okuryazarlığı testinin 11 maddesinden alınan veriler 506 kişilik Türkiye çalışma grubu için MTK'nın varsayımları olan tek boyutluluk, monotonluk ve yerel bağımsızlık varsayımlarını karşılamıştır. MTK ve Karma MTK modelleri ile yapılan kestirimler sonucunda PISA 2015 fen maddelerinin iki gizil sınıflı Karma 2 PLM'ye uygun olduğu bulunmuştur. 2 Gizil Sınıflı Karma 2PLM model temel alınarak analizler gerçekleştirilmiş ve yanıtlayıcıların yaklaşık %91'nin birinci gizil sınıfta, %9'unun ise ikinci gizil sınıfta yer aldığı sonucuna ulaşılmıştır. Sınıflama oranına ilişkin Kernel yoğunluk grafiği model veri uyumunun bir göstergesi olarak belirlenmiş veri setinin gizil sınıf ayrımının 2 Gizil Sınıflı Karma 2 PLM'ye göre model veri uyumu sınıflama oranına ilişkin Kernel yoğunluk grafiği incelenmiştir. Kernel yoğunluk grafiği, model veri uyumunun bir göstergesi olarak ele alınmış ve veri setinin 2 Gizil Sınıflı Karma 2PLM'ye göre model veri uyumunu sağladığı bulunmuştur.

Araştırmanın nihai amacı olan PISA 2015 Türkiye fen okuryazarlığı maddelerinin gizil sınıf yöntemiyle DMF içerip içermediğine bakılmıştır. Madde 5, 6, 9 ve 11'in tek biçimli, madde 4'ün ise hem tek biçimli hem de tek biçimli olmayan DMF gösterdiği bulunmuştur. Bu sonuçla beraber PISA 2015 fen okuryazarlığı maddelerinden incelenen 11 maddenin 5'inin DMF içerdiği sonucuna ulaşılmıştır.

Bu çalışmada "cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma, okula ait hissetme ve okuma becerisi" gözlenen değişkenlerine göre DMF kaynağı olarak belirlenmiş olup, bunların gizil sınıfları açıklamada yetersiz kaldığı sonucuna ulaşılmıştır. Bu durum, bireylerin fen okuryazarlığı maddelerine ilişkin tepkilerinde bu değişkenlerden farklı bir boyutun etkisi ile gizil sınıflara ayrıldıklarını göstermektedir. Cohen ve Bolt, cinsiyete göre DMF

analizi yaptıkları çalışmalarında cinsiyetin bir DMF kaynağı olduğunu belirtmişlerdir. Ancak gizil sınıf analizi sonucunda bilişsel bir özelliğin başka bir DMF kaynağı olduğunu ve sadece gözlenen grup değişkenine göre yapılan DMF analizlerinin DMF'nin asıl kaynağını belirlemede yetersiz kaldığını belirtmişlerdir. Alan yazınında DMF'nin kaynağını belirlemede karma modellere dayalı yöntemlerin daha iyi sonuçlar verdiği de ortaya konmuştur. Sonuç olarak gözlenen değişkene göre DMF kaynağı olarak belirlenen bir değişkenin gizil sınıflara ilişkin DMF'nin asıl kaynağının gözden kaçırılmasına neden olabileceği sonucuna varılmıştır.

“Cinsiyet, bilgi ve teknoloji iletişim kaynakları, fen öğretiminden zevk alma, okula ait hissetme” değişkenleri modele ayrı ayrı kovaryant değişken olarak dâhil edildiğinde bireylerin sınıflara atanma doğruluğu yükselmiştir. Dolayısıyla bu çalışmada kovaryant değişkenin gizil değişkenin bir yordayıcısı ya da önsel bilgi kaynağı olarak görev yaptığı sonucuna ulaşılmıştır.

Bu çalışmada, gözlenen değişkene dayalı DMF kaynağı olan değişkenlerin karma MTK modele dahil edildiğinde gizil sınıflarla örtüşme oranının oldukça düşük olduğu gözlenmiştir. Yine bu değişkenler karma MTK modele dahil edildiğinde bireylerin yetenek parametre kestiriminde anlamlı düzeyde bir değişime neden olmamışlardır. Ölçme işlemlerinde nihai amaç bireylerin gerçek yeteneklerinin en az hata ile kestirilmesidir. Dolayısıyla bir değişken her ne kadar gözlenen değişkene dayalı DMF kaynağı olsa da bu değişkenler gerçek yeteneğin belirlenmesinde anlamlı bulunmamıştır. Bu çalışma gözlenen değişkene dayalı DMF kaynağının gizil değişken ile zayıf düzeyde ilişkisi olması durumunda, test sonuçlarının geçerliğini önemli ölçüde etkilemediğini göstermiştir.