İçtimaiyat
Sosyal Bilimler Dergisi | Journal of Social Sciences
DergiPark AKADEMİK doi Foundation akademik intihal.net E-ISSN 2602-3377

Investigation of Classification Validity in TIMSS 2019 Proficiency Classification of Students in Terms of Various Variables

Yazar(lar) / Author(s)
Dr, Esra Oyar
Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye
e-posta: esratas@gazi.edu.tr.
(Sorumlu Yazar/Corresponding author)

Makale Bilgileri/Article İnformation
Tür-Type: Araştırma makalesi-Research article
Geliş tarihi-Date of submission: 15. 07. 2024
Kabul tarihi-Date of acceptance: 26. 10. 2024
Yayım tarihi-Date of publication: 31. 11. 2024

Hakemlik-Review
Hakem sayısı-Reviewers: İki Dış Hakem-Two External
Değerlendirme-Rewiev: Çift Taraflı Kör Hakemlik-Double-blind

Etik/Ethics
Etik beyan- Ethical statement: Bu çalışmanın hazırlanma sürecinde etik ilkelere uyulmuştur.
Benzerlik taraması- Plagiarism checks Yapıldı-İntihal.net-Yes-İntihal.net
Etik bildirim- Complaints
ictimaiyatdergi@gmail.com

Çıkar çatışması-Conflict of interest
Çıkar çatışması beyan edilmemiştir.
The Author(s) declare(s) that there is no conflict of interest

Finansman-Grant Support
Herhangi bir fon, hibe veya başka bir destek alınmamıştır.
No funds, grants, or other support was received.

Abstract

In this study, it was aimed to determine the variables affecting students' proficiency classification by using the data of 4th-grade students participating in the TIMSS 2019 application in the fields of mathematics and science. For this purpose, it was tried to provide evidence for classification validity with the variables of school belonging, bullying, home resources for learning, self-efficacy for computer use, disorderly behavior in Math lessons, like learning Math/Science, confident in Math/Science and instructional clarity in Math/Science lessons. The study was conducted with a correlational design. The sample of the study consisted of 3887 students in both lessons, which remained as a result of the missing data deletion and assignment processes from 4028 students who originally participated in the application. Logistic regression and discriminant analysis were used to analyze the data. As a result of the study, it was determined that 41.6% for Mathematics and 43% for Science in logistic regression analysis and 42.5% for Mathematics and 45% for Science in discriminant analysis were correctly classified through independent variables. The results obtained from the study were discussed in the light of the literature and recommendations for both researchers and practitioners were presented.

Keywords: Logistic regression, discrimination analysis, TIMSS 2019, Mathematics and Science, international benchmarks.

Öğrencilerin TIMSS 2019 Yeterlik Sınıflamasında Sınıflama Geçerliğinin Çeşitli Değişkenler Açısından İncelenmesi

Öz

Bu çalışmada, TIMSS 2019 uygulamasına katılan 4. sınıf öğrencilerinin matematik ve fen bilimleri alanlarındaki verileri kullanılarak öğrencilerin yeterlik sınıflandırmasına etki eden değişkenlerin belirlenmesi amaçlanmıştır. Bu amaçla okula aidiyet, zorbalık, öğrenme için ev kaynakları, bilgisayar kullanımı için öz yeterlik, Matematik derslerinde düzensiz davranış, Matematik/Fen öğrenmeyi sevme, Matematik/Fen'de kendine güven ve Matematik/Fen derslerinde öğretimsel netlik değişkenleri ile sınıflama geçerliğine kanıt sağlanmaya çalışılmıştır. Çalışma korelasyonel bir tasarımla yürütülmüştür. Çalışmanın örneklemi, başlangıçta uygulamaya katılan 4028 öğrenciden kayıp veri silme ve atama işlemleri sonucunda kalan her iki dersteki 3887 öğrenciden oluşmaktadır. Verilerin analizinde lojistik regresyon ve diskriminant analizi kullanılmıştır. Çalışma sonucunda, öğrencilerin yeterlilik sınıfları bağımsız değişkenler olarak ele alındığında, lojistik regresyon analizinde Matematik için %41,6 ve Fen Bilimleri için %43; diskriminant analizinde ise Matematik için %42,5 ve Fen Bilimleri için %45 olarak belirlenmiştir. Çalışmadan elde edilen sonuçlar literatür ışığında tartışılmış ve hem araştırmacılara hem de uygulayıcılara yönelik öneriler sunulmuştur.

Anahtar Kelimeler: Lojistik regresyon, ayırma analizi, TIMSS 2019, Matematik ve Fen Bilimleri, yeterlik sınıflamaları.

Atıf- Citation (APA)
Oyar, E. (2024). Investigation of classification validity in TIMSS 2019 proficiency classification of students in terms of various variables. *İçtimaiyat*, Cilt 8, Türk Eğitim Sisteminde Değişimler ve Yeni Eğilimler Özel Sayısı, ss. 57-76. DOI: https://doi.org/10.33709/ictimaiyat.1515877

1. Introduction

There are different definitions of validity in the literature. In its most general definition, validity is the degree to which a measurement tool can accurately measure the characteristic it aims to measure without confusing it with any other characteristic (Tekin, 1977). In other words, it is the degree to which the measurement tool serves the purpose (Crocker & Algina, 2006). Kane (2001) determined a criterion as the value of the trait of interest and stated that the test would be valid if an accurate prediction is made according to this criterion. Cureton (1951) stated that validity can be accurately determined with criterion-resistant models. In addition, Campbell and Fiske (1959) stated that there are convergent and discriminant validity under criterion validity. However, criterion validity may not be sufficient for all tests. In this case, another strategy that can be considered is content validity, and especially in achievement tests, evidence of content validity needs to be provided (Schmidt, 2012). In addition to these definitions, a concept that has been introduced is construct validity. Cronbach and Meehl (1955) stated that construct validity would be an alternative to content and criterion validity. Messick (1995) stated that with the unified validity concept, all types of validity can be gathered under the same roof of construct validity. In addition to the types of validity defined in this section, it is also suggested to use the concept of "classification validity" to determine the consistency of classification decisions made with measurement tools used for selection, placement and diagnosis (Erkuş, 2004). It is stated that the main focus in the concept of validity, which is examined under different subheadings, is that a test should have a high predictive power of success and failure and contribute to making the right decisions (Murphy & Davidshofer, 2001). At this point, the concept of classification validity is not separated from other concepts (Saral, 2012); however, while construct and criterion validity are more related to the structure of the questions in the test and their relationship with the criterion, classification validity is related to whether the decisions made based on the test results are correct (Güzeller & Kelecioğlu, 2006).

There are different statistical methods used to determine the accuracy in classifying individuals in order to collect evidence for classification validity. When the literature is examined, it is seen that there are studies stating that logistic regression and discriminant analysis provide evidence for classification validity (Atar, 2012; Güzeller & Kelecioğlu, 2006; Kan, 2004; Saral, 2012; Taşdemir, 2015). Güzeller and Kelecioğlu (2006) examined the validity of placements based on subtest raw scores in Secondary Education Institutions Student Selection Exams. For this purpose, discriminant analysis was used as the analysis method. The discrimination functions obtained based on the subtests were found to be effective in separating public science high schools from private and Anatolian high schools; however, they were not effective in separating students placed in private science high schools and Anatolian high schools. The correct classification rate of public science high schools was 96%; the correct classification rate of private science high schools was 36.7% and the correct classification rate of Anatolian high schools was 52.7%. In the study, Atar (2012) examined the classification accuracy to determine whether the calculation method used by ÖSYM in placing students into teaching programs requiring special aptitude works in the expected direction in real practice. For this purpose, logistic regression and discriminant analysis were performed. As a result of the study, it was seen that the sub-score type weights determined by ÖSYM did not work as expected in practice and an alternative calculation method was presented. In this study, two methods recommended in the literature were used. These methods are logistic regression and discriminant analysis. In both analysis methods, the class of individuals is predicted from the independent variables determined. At this point, it is important to determine the independent variables that are thought to be effective in assigning individuals to the proficiency class. In this direction, a literature review was conducted and the variables to be included in the study were determined.

There are many studies examining the variables affecting students' achievement in Mathematics and Science in large-scale exams conducted in Türkiye (Akyüz, 2006; Coşkun & Karakaya-Özer, 2023; Karabay, Yıldırım, & Güler, 2015; Karalı, Varol-Palancıoğlu, & Aydemir, 2022; Özer & Anıl, 2011; Şahin, Çelik, & Yıldırım, 2022; Şevgin & Eranıl, 2023; Yavuz, Demirtaşlı, Yalçın, & Dibek, 2017). When the studies conducted by years are examined, Akyüz (2006) examined the variables affecting TIMSS mathematics achievement with HLM (Hierarchical Linear Modeling) analysis and determined that students' home resource status affected achievement. Doğan and Barış (2010) examined the predictive power of attitude, value and self-efficacy variables on students' Mathematics achievement in TIMSS 1999 and TIMSS 2007 assessments by using standard multiple regression technique and found that all three variables were significant in predicting Mathematics achievement in TIMSS 2007. Yavuz et al. (2017) examined the variables affecting students' Mathematics achievement with HLM analysis and found that while the variable of students' self-confidence in mathematics had a significant effect on TIMSS achievement in 2011, it did not have a significant effect in TIMSS 2007. They also stated that the value that students placed on mathematics did not show a significant relationship with students' mathematics achievement in both implementation periods. Şahin et al. (2022) used path analysis in their study in which they examined various factors affecting the Mathematics achievement of 8th grade students participating in the TIMSS 2019 application and as a result of the study, they determined that instructional clarity significantly affected academic achievement. In another study, Çoşkun and Karakaya-Özyer (2023) examined the factors affecting the Mathematics achievement of 8th grade students in TIMSS 2011, 2015 and 2019 applications with HLM analysis. As a result of the study, it was determined that students' self-confidence was one of the most important characteristics affecting their Mathematics achievement in all three applications; in addition, home resources for learning also affected students' Mathematics achievement. In their study, Şevgin and Eranıl (2023) examined the variables in both Mathematics and Science that affect school engagement in the TIMSS 2019 application and as a result of the study, it was determined that the order of importance of the variables in both fields was similar and, in this context, one of the most important variables affecting school engagement was the level of bullying at school. Considering the positive effect of school engagement on students' academic achievement (Ladd & Dinella, 2009; Upadyaya, & Salmela-Aro, 2013), it is considered appropriate to determine the effect of school bullying variable on the classification of students into international benchmarks in Mathematics and Science. When the studies are examined, it is seen that the factors affecting achievement are handled in various ways. In this study, school belonging, school bullying, home resources for learning, self-efficacy for computer use, which are thought to affect students' proficiency classification, were included in the analyses for both fields. In addition, the variables of disorderly behaviour in Math lessons, like learning Math, confident in Math and instructional clarity in Math lessons were included in the analysis for Mathematics, and the variables of like learning Science, confident in Science and instructional clarity in Science lessons were included in the analysis for Science.

The main purpose of TIMSS is to help improve education and training in mathematics and science worldwide. Considering the purpose of TIMSS, this test, which is also administered in Türkiye, is used to monitor student achievement trends and to compare achievement results with other countries participating in the test. In achievement comparisons, raw scores can be compared or proficiency classifications can be used. In this case, it is thought to be important to determine the effective variables in the formation of proficiency classifications and to provide direction for future studies.

The following research problems were formulated within the scope of the study:

- How is the classification accuracy according to logistic regression analysis in terms of various variables in the classification of 4th grade students participating in the TIMSS 2019 into international benchmarks in Mathematics and Science?
- How is the classification accuracy according to the discriminant analysis in terms of various variables in the classification of 4th grade students participating in the TIMSS 2019 into international benchmarks in Mathematics and Science?
- To what extent do the classification validity evidence obtained from logistic regression and discriminant analysis overlap?

2. Methods

In this section, the methodology of the study, the population and sample, the dependent and independent variables and the analysis of the data were presented.

2.1. Study Design

In this study, it was aimed to examine and compare the results of different analysis techniques through independent variables in the classification of students according to their proficiency levels in Mathematics and Science with TIMSS 2019 data. A correlational research model was used in the study. The aim of the correlational research model is to determine the existence and degree of the relationship between variables and to predict the other from the data of one variable (Karasar, 2014).

2.2. Study Sample

The data used in this study consist of student data participating in TIMSS 2019. In TIMSS 2019, a two-stage process was followed in sample selection. First, random sampling method was used in which all schools in Türkiye have equal probability of being selected. In the second stage, a random branch was selected from each selected school and the exam is administered. As a result of this sampling, a total of 4028 4th grade students participating in the TIMSS 2019 application constitute the sample of the study.

2.3. Variables Included in the Analysis

In this section, the dependent and independent variables and the measurement tools used to measure these variables are discussed.

2.3.1. Dependent Variable

In TIMSS 2019, international mathematics and science benchmarks are defined for students participating in the application at the 4th grade level. These levels are briefly as follows:

1 = Below 400

2 = At or above 400 but below 475

3 = At or above 475 but below 550

4 = At or above 550 but below 625

5 = At or above 625

Within the scope of the study, both Mathematics and Science classifications were included in the analysis as separate dependent variables.

2.3.2. Independent Variables

Sense of School Belonging: It is administered to determine students' sense of belonging to school. Within the scope of this measurement tool, 5 items are directed to the students and the students are asked to respond to these items with a 4-point Likert scale with 1: strongly agree and 4: strongly disagree. The reliability coefficient (Cronbach α) for the scale was estimated as 0.66 (Yin & Fishbein, 2019).

Students' Bullying: This instrument asks students how often they are subjected to disturbing behaviors by other students. There are 11 items in the instrument and students are asked to respond to the items on a 4-point Likert scale: 1: at least once a week; 2: once or twice a month; 3: a few times a year; and 4: never. The reliability coefficient (Cronbach α) for the scale was estimated as 0.83 (Yin & Fishbein, 2019).

Home resources for learning: In this context, students are asked about the educational resources available in their homes. While 5 of these items are the same for all students, 4 of them include country-specific home resources. Students are asked to answer yes/no whether they have these resources or not. The reliability coefficient (Cronbach α) for the scale was estimated as 0.75 (Yin & Fishbein, 2019).

Self-Efficacy for Computer Use: In this context, students are asked about the extent to which they feel competent in using computers. There are 7 items in total in this measurement tool and students are asked to respond to these items with a 4-point Likert scale as "never", "some lessons", "about half the lessons" and "ever tired almost every lesson".

Disorderly Behavior in Math lessons: In this context, there are items to determine the frequency of disruptive behaviors in Mathematics lessons. There are a total of 6 items in this measurement tool and students are asked to respond to these items with a 4-point Likert scale with 1: strongly agree and 4: strongly disagree. The reliability coefficient (Cronbach α) for the scale was estimated as 0.83 (Yin & Fishbein, 2019).

Like Learning Mathematics/Science: In this context, items were presented to the students regarding their liking for the related course in both Mathematics and Science. There are 9 items in both measurement tools and students are asked to respond to the items on a 4-point Likert scale with 1: strongly agree and 4: strongly disagree. The reliability coefficient (Cronbach α) for the scale related to like learning mathematics and science were estimated as 0.88 and 0.86, respectively (Yin & Fishbein, 2019).

Confidence in Mathematics/Science: In this context, items were presented to the students to measure their confidence in the relevant course, both in Mathematics and Science. There are 7 items in both measurement tools and students are asked to respond to the items on a 4-point Likert scale with 1: strongly agree and 4: strongly disagree. The reliability coefficient (Cronbach α) for the scale related to confidence in mathematics and science were estimated as 0.84 and 0.81, respectively (Yin & Fishbein, 2019).

Instructional Clarity for Mathematics/Science: In this context, students were presented with items to measure their confidence in both Mathematics and Science. There are 6 items in both measurement tools and students are asked to respond to the items on a 4-point Likert scale with 1: strongly agree and 4: strongly disagree. The reliability coefficient (Cronbach α) for the scale related to instructional clarity for mathematics and science were estimated as 0.70 and 0.76, respectively (Yin & Fishbein, 2019).

2.4. Data Analysis

2.4.1. Preliminary Analysis

Two different methods were used to analyze the data. These were logistic regression analysis and discriminant analysis. Below were the equations and explanations of these two methods. However, some assumptions needed to be tested before these methods. Some assumptions needed to be tested for multivariate analysis before the two methods used in the analysis. In this section, these assumptions were given first.

Missing values that exhibit a random pattern in large data sets do not cause serious problems in the analysis (Tabachnick & Fidell, 2013). It is also stated that if the amount of missing data does not exceed 5% of the total data, it can be ignored. Descriptive statistics on the percentages of missing data were presented in Table 1.

Table 1: Missing Values

Variables	f	%
School_Belong	135	3,35%
Bullying	126	3,12%
Home_resource	279	6,9%
Self_efficacy_ICT	23	0,57%
Ins_Clarity_M	72	1,78%
Liking_M	27	0,67%
Confident _M	83	2,06%
Dis_Beh_M	72	1,78%
Ins_Clarity_S	46	1,14%
Liking_S	44	1,09%
Confident _S	61	1,51%

School_Belong: Sense of School Belonging; Bullying: Students' Bullying; Home_resource: Home resources for learning; Self_efficacy_ICT: Self-Efficacy for Computer Use; Ins_Clarity_M: Instructional Clarity for Mathematics; Liking_M: Like Learning Mathematics; Confident _M: Confidence in Mathematics; Dis_Beh_M: Disorderly Behavior in Math lessons; Ins_Clarity_S: Instructional Clarity for Science; Liking_S: Like Learning Science; Confident _S: Confidence in Science.

Table 1 showed that the missing data rates for almost all variables were below 5%. Only for the home resources variable, this value was more than 5%. When the literature is examined, many methods are suggested for the missing data problem (Tabachnick & Fidell, 2013). In this study, the average assignment method was used for missing data and average values were given for missing values of individuals.

Outliers are extreme values that are not considered appropriate for the data set when compared to other data. These extreme values may be erroneous or may reflect reality. The z scores are calculated for the data related to the variables in the analysis and the data that are outside the ±3 values can be characterized as outliers (Çokluk, Şekercioğlu & Büyüköztürk, 2012). For the outlier analysis, z scores were obtained and a total of 141 student data were excluded from the analysis for each variable. As a result, 3887 student data were analyzed.

For normality, skewness and kurtosis values of the data are calculated and if these values are within the range of ±1, it is shown as evidence that the data do not deviate excessively from normal (Çokluk et al., 2012). Statistics regarding the normality of the data were calculated and skewness and kurtosis values were examined for each variable. As a result of the examinations, it was determined that the data showed normal distribution.

2.4.2. Logistic Regression

Logistic regression analysis is an analysis method that allows the prediction of group membership from a group of continuous, binary or a mixture of these variables. In a linear regression analysis, it is not correct to include a categorical variable in the analysis as continuous (Tabachnick & Fidell, 2013). In this case, logistic regression analysis is used because the variables are categorical or ordinal. The basis of logistic regression analysis is to create a regression equation that will be used to predict which group individuals are in, in other words, to predict group membership (Çokluk, et al., 2012)

Logistic regression analysis is analyzed under 3 types. If the dependent variable consists of three or more categories and these categories are ordinal, it is called Ordinal Logistic Regression. In this study, there are 5 ordinal categories (advanced level, upper level, middle level, lower level and below lower level). Therefore, the study can be expressed as Ordinal Logistic Regression analysis. Independent variables can be categorical or continuous.

In ordered logistic regression, threshold values are obtained as 1 minus the number of classes in the dependent variable and are called "C". For example, for c=1, these values represent the threshold value between class 1 and class 2 in the dependent variable. That is, when the effect of the predictor variables is held constant or zero, it is the estimated cut-off point used to distinguish the membership of class 1 in the dependent variable from the other classes.

The mathematical basis of logistic regression is probability, odds and logarithm of odds.

$$odds = \frac{p(X)}{1-\mathrm{p(X)}} \qquad (1)$$

Here p(X) gives the probability of an event happening and 1-p(X) gives the probability of the event not happening. The outcome model obtained here is a non-linear function. The main focus of logistic regression is the concept of "logit". The logit concept is equal to the natural logarithm of the odds value. The result is given in equation (2).

$$\widehat{Y_\iota} = \frac{e^u}{1+e^u} \qquad (2)$$

Y_i : The estimated probability that person i is in any category of the dependent variable

u: $B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \ldots\ldots\ldots\ldots + B_k X_k$

As a result, the logistic regression formula we obtained can be expressed in Equation (3).

$$In\,(\frac{\hat{Y}}{1-\hat{Y}}) = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \ldots\ldots\ldots\ldots + B_k X_k \qquad (3)$$

As can be seen in Equation (3), there are k independent variables in the logistic regression model (k=1,2,3,….,k). These variables can be included in the model in different ways. For this reason, the model to be used should be decided first. In this study, the standard (enter) method was used to include the variables in the analysis. In this method, all independent variables are included in the regression model as a block and parameter estimates are calculated for each block (Çokluk et al., 2012).

2.4.3. Testing Assumptions in Logistic Regression Analysis

In logistic regression analysis, there are no assumptions regarding the distribution of the independent variables in the model (Tabachnick & Fidell, 2013). However, as in other analyses, multicollinearity and outlier were checked before the analysis.

2.4.4. Discrimination Analysis

One of the methods that can be used in a study with a dependent variable consisting of two or more categories is discriminant analysis. In discriminant analysis, the relationships between variables can be determined as well as which of the variables contributes best to the classification (Tabachnick & Fidell, 2013).

In discriminant analysis, individuals are already members of a group, but it should be taken into account that some individuals are incorrectly assigned to groups (Çokluk et al., 2012). In other words, discriminant analysis provides evidence for the validity of the classification made at this point.

In the discriminant analysis, each individual should have a score or scores belonging to one or more quantitative variables and a categorical variable value indicating group membership. In this analysis, quantitative variables are called "independent variables" and variables indicating group membership are called "dependent variables".

A function calculated in the discriminant analysis is the discriminant function, and the discriminant function is obtained as the number of degrees of freedom for the groups or the number of predictor variables, whichever is smaller. In analyses, one or two discriminant functions usually account for a significant proportion of the discriminant power. Discriminant functions are similar to regression equations.

$$D_i = a_i + b_{i1}X_1 + b_{i2}X_2 + \cdots + b_{ip}X_p \qquad (4)$$

D_i : predicted score for i. discriminant function

X_p: raw score for the predictor variable of p

a_i: constant for discriminant function i

b_{ip}: the unstandardized partition function coefficient of the p. predictor variable.

A classification equation is obtained for each group to assign individuals to groups (Tabachnick and Fidell, 2013). The individual is assigned to the group with the highest classification score.

$$C_j = c_{j0} + c_{j1}X_1 + c_{j2}X_2 + \cdots + c_{jp}X_p \qquad (5)$$

C_j : classification function score for group j

X_p : p. raw score of predictor variable of p.

c_{jp} : classification function coefficient of the predictor variable of p

c_{j0} : the classification function constant for group j.

2.4.5. Testing Assumptions in Discriminant Analysis

First of all, it should be checked whether the assumptions of the discriminant analysis are met in order for the analysis to be in the most appropriate way and to minimize misclassification. As in multivariate statistical analyses, sample size, normal distribution, homogeneity of variance-covariance matrices, outliers and multicollinearity are the assumptions to be examined in discriminant analysis. Regarding the sample size in discriminant analysis, it is stated that the sample should be large enough (e.i, at least 20 for each group) (Göçer-Sahin, 2022). Since TIMSS 2019 data were used in the study, the sample is large enough and this assumption was met. In order to meet the assumption of normal distribution, skewness and kurtosis values were checked. Since the skewness and kurtosis values of the variables varied between ±1, it was accepted that the normality assumption was met. Box's M statistic results were analyzed to test the assumption of equal variance-covariance. A significant Box's M statistic means that the variance-covariance matrices are not homogeneous. If the result is not significant, quadratic decomposition analysis is used. According to the result obtained, it is seen that the covariance matrices were not equal to each other. The discrimination power of the linear discriminant function is greatly affected when the covariance matrices are not equal. Therefore, it was decided to use quadratic discriminant analysis in the analysis due to the unequal covariance matrices. At the beginning of the analysis, necessary checks were made for missing data and outliers and the data were organized. The problem of multicollinearity occurs when there is a correlation above .90 between variables (Tabachnick & Fidell, 2013). In order to check this assumption, the correlation between the total scores of the students from each independent variable was examined and it was seen that the highest correlation value was not high enough to create a multicollinearity problem.

3. Findings

3.1. Findings for Logistic Regression

First, logistic regression analysis was performed to determine the classification accuracy in terms of various variables in the classification of 4th grade students participating in TIMSS 2019 according to international benchmarks in Mathematics and Science. The fit values for the entire model were given in Table 2.

Table 2. Results of Fit Values for the Model

	Model	-2 Log likelihood (-2LL)	Chi-Square	df	Sig.
4th-grade Mathematics	Intercept Only	12033.223			
	Final	9998.329	2034.895	8	.000
4th-grade Science	Intercept Only	11253.808			
	Final	9610.264	1643.544	7	.000

The log-likelihood values in the results are related to the fit of each model to the data. If the test results are significant, it means that the variables included in the model contribute significantly to the improvement of the model. As a result, all independent variables included in each model contributed significantly to the identification of the models. The first row in the table (intercept only) shows the results for the model without independent variables, while the second row shows the results when independent variables are included in the model. When the values were examined, it was seen that the relevant value was significant. In this case, it was concluded that the model fits the data well for both subject areas.

The coefficients of the variables in the logistic regression model for the classification of students according to their proficiency levels, the standard errors of these coefficients, "Wald Statistics" and the significance levels of these statistics were presented in Table 3.

Table 3. Parameter Estimates

	Variables		B	Std. error	Wald	df	Sig.
4th-grade Mathematics	Threshold	C=1	8.264	.327	636.922	1	.000
		C=2	9.755	.335	848.267	1	.000
		C=3	11.440	.348	1082.247	1	.000
		C=4	13.485	.365	1364.940	1	.000
	School_Belong		-.035	.018	3.685	1	.055
	Bullying		.091	.018	27.190	1	.000
	Home_resource		.629	.020	954.857	1	.000
	Self_efficacy_ICT		.047	.017	7.517	1	.006
	Ins_Clarity_M		.091	.019	23.192	1	.000
	Liking_M		-.045	.022	4.240	1	.039
	Confident _M		.398	.020	388.577	1	.000
	Dis_Beh_M		.005	.017	.106	1	.745
4th-grade Science	Threshold	C=1	7.102	.308	531.157	1	.000
		C=2	8.584	.315	742.633	1	.000
		C=3	10.451	.329	1009.050	1	.000
		C=4	12.865	.348	1368.259	1	.000
	School_Belong		-.045	.018	6.609	1	.010
	Bullying		.094	.017	31.018	1	.000
	Home_resource		.600	.020	894.804	1	.000
	Self_efficacy_ICT		.093	.017	28.799	1	.000
	Ins_Clarity_S		.037	.020	3.284	1	.070
	Liking_S		.084	.020	17.292	1	.000
	Confident _S		.205	.021	96.693	1	.000

In table 3, when the threshold parameters for classification in mathematics were examined, the threshold value between 1st and 2nd grade in the dependent variable was estimated as 8.26. That is, when the effect of predictor variables is zero, the estimated cut-off point obtained for the 1st class membership in the dependent variable is 8.26. This value was estimated as 9.76 for the separation of 2nd and 3rd class, 11.44 for the separation of 3rd and 4th class and 13.49 for the separation of 4th and 5th class. In science, the threshold value between 1st and 2nd grade in the dependent variable was estimated as 7.10. This value was estimated as 8.58 for the separation of 2nd and 3rd class, 10.45 for the separation of 3rd and 4th class and 12.87 for the separation of 4th and 5th class.

The beta values in the table are interpreted as regression coefficients. For example, for Mathematics, when the effect of other independent variables was held constant, an increase of 1 unit in the variable “home resources for learning” corresponded to an increase of 0.629 units in the logarithm of the odds ratio. Similarly, in Science, holding the effect of other independent variables constant, a 1-unit increase in the variable “home resources for learning” corresponded to a 0.600-unit increase in the logarithm of the likelihood ratio. These interpretations were also valid for the other variables.

When the Wald statistic values in the table were examined, the effect of the variables "disorderly behaviour in Math lessons" and "school belonging" on the classification was found to be statistically insignificant for Mathematics ($p>0.05$). In the field of science, only the effect of "instructional clarity" variable on classification was found to be statistically insignificant ($p>0.05$). Apart from this, the other variables included in the study made a significant contribution to classification.

Pseudo R^2 values obtained as a result of the analysis show the power of the independent variables in classifying the model correctly. Findings regarding the values were given in Table 4.

Table 4. Pseudo R^2 Values

	Cox and Snell	Nagelkerke
4th-grade Mathematics	0.408	0.427
4th-grade Science	0.345	0.364

The interpretation of the values in the table is similar to the R^2 interpretations in regression analysis. When the table values were examined, it could be said that the independent variables have an explanatory power of 41% in the classification of students according to the Cox and Snell R^2 value for mathematics. In the field of science, it was seen that the independent variables had an explanatory power of approximately 35% in the classification of students. On the other hand, Nagelkerke R^2 values, which is a modification of the Cox and Snell R^2 value, are also interpreted because it cannot approach 1 and therefore creates difficulty in interpretation. This value was estimated to be approximately 43% for mathematics and 36% for science.

The classification performance of logistic regression analysis according to students' TIMSS 2019 Grade 4 mathematics proficiency classification were presented in Table 5.

Table 5. Correct Classification Percentages (Mathematics)

Observed		Estimated				
		1	2	3	4	5
1	f	**138**	122	183	12	1
	%	**30.3%**	26.8%	40.1%	2.6%	0.2%
2	f	86	**121**	373	95	4
	%	12.7%	**17.8%**	54.9%	14.0%	0.6%
3	f	45	105	**611**	310	35
	%	4.1%	9.5%	**55.2%**	28.0%	3.2%
4	f	9	43	373	**546**	113
	%	0.8%	4.0%	34.4%	**50.4%**	10.4%
5	f	1	4	69	286	**202**

	%	0.2%	0.7%	12.3%	50.9%	**35.9%**

When the table values were examined, the variables included in the analysis correctly classified 30.3% of the students in the first proficiency class, while the correct classification of the students in the other proficiency classes was 17.8%, 55.2%, 50.4% and 35.9%, respectively. According to the results of the analysis, the number of correctly classified students was 1618 (41.6%).

The performance of logistic regression analysis in classifying students according to TIMSS 2019 Grade 4 science proficiency classification were presented in Table 6.

Table 6. Correct Classification Percentages (Science)

Observed		Estimated				
		1	2	3	4	5
1	f	**99**	75	159	14	0
	%	**28.5%**	21.6%	45.8%	4.0%	0.0%
2	f	50	**61**	359	100	0
	%	8.8%	**10.7%**	63.0%	17.5%	0.0%
3	f	29	61	**670**	476	8
	%	2.3%	4.9%	**53.9%**	38.3%	0.6%
4	f	8	14	460	**800**	42
	%	0.6%	1.1%	34.7%	**60.4%**	3.2%
5	f	1	1	62	293	**45**
	%	0.2%	0.2%	15.4%	72.9%	**11.2%**

When the table values were examined, the variables included in the analysis correctly classified 28.5% of the students in the first proficiency class, while the correct classification of the students in the other proficiency classes was 10.7%, 53.9%, 60.4% and 11.2%, respectively. According to the results of the analysis, the number of correctly classified students was calculated as 1675 (43%).

3.2. Findings for Discriminant Analysis

Secondly, a discriminant analysis was conducted to determine the classification accuracy in terms of variables in the classification of 4th grade students participating in TIMSS 2019. First, the eigenvalue table was examined to determine the significance of the discriminant functions. The results were presented in Table 7.

Table 7. Eigenvalues

	Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
4th-grade Mathematics	1	.694	92.3	92.3	.640
	2	.053	7.0	99.3	.224
	3	.004	.5	99.8	.061
	4	.001	.2	100.0	.037
4th-grade Science	1	.522	88.7	88.7	.586
	2	.062	10.6	99.3	.242
	3	.004	.7	100.0	.063

4	.000	.0	100.0	.005

When the table values were examined, the first discriminant function with an eigenvalue of 0.694 was used to explain the inter-group variability for Mathematics. Although an exact value was not accepted, eigenvalues greater than 0.40 were considered good. According to the results, only one eigenvalue exceeded the specified limit. When the total variance explanation ratios were analyzed, the first function for Mathematics explained 92.3% of the total variance. When canonical correlation values were analyzed, it was seen that the highest value was in the first discriminant function. In science, the first discriminant function was used with an eigenvalue of 0.52 to explain the variability between groups. The first function explained 88.7% of the total variance for the field of Science and the highest value was in the first discriminant function.

Wilk's Lambda and F statistics of the variables were calculated to test whether the selected variables separated the groups significantly. Thus, it was tested whether there was a significant difference in the five competency classes according to the independent variables. The findings were given in Table 8.

Table 8. Wilks' Lambda Test for Equality of Group Means

		Wilks' Lambda	F	df1	df2	Sig.
4th-grade Mathematics	School_Belong	.976	23.620	4	3882	.000
	Bullying	.958	42.681	4	3882	.000
	Home_resource	.720	377.389	4	3882	.000
	Self_efficacy_ICT	.945	56.292	4	3882	.000
	Dis_Beh_M	.981	19.251	4	3882	.000
	Ins_Clarity_M	.918	86.138	4	3882	.000
	Liking_M	.940	61.634	4	3882	.000
	Confident _M	.809	228.646	4	3882	.000
4th-grade Science	School_Belong	.975	24.61	4	3882	.000
	Bullying	.963	37.58	4	3882	.000
	Home_resource	.725	367.54	4	3882	.000
	Self_efficacy_ICT	.931	71.42	4	3882	.000
	Ins_Clarity_S	.926	77.62	4	3882	.000
	Liking_S	.896	112.85	4	3882	.000
	Confident _S	.887	123.15	4	3882	.000

In Table 8, it was seen that the differences between the groups in student scores for all independent variables were significant. If the Wilk's Lambda values obtained as a result of the analysis are close to 1, it shows that the effect of subtests in separating groups is not very high. In this case, when the values were analyzed, it could be said that " home resources for learning " and "student confident in math" variables are more effective than other variables in mathematics, and "home resources for learning " and "student confident in science" and "students like learning science" variables in science.

To decide which independent variable is the most effective in discrimination, it is necessary to examine the standardized discriminant function and the structural matrix that reduces it. The standardized coefficients for the discriminant functions obtained for the classification of students for Mathematics were presented in Table 9.

Table 9. Standardized Coefficients for Discriminant Functions (Mathematics)

	Function			
	1	2	3	4
Home_resource	.746*	.210	-.112	.151
Confident _S	.572*	-.401	.259	-.176
Self_efficacy_ICT	.257	.474*	.250	-.381
School_Belong	.146	.419*	-.062	-.382
Bullying	.242	.247*	.054	.091
Ins_Clarity_S	.343	.345	-.500*	-.166
Liking_S	.289	.298	.484*	.117
Dis_Beh	.165	.086	-.158	.613*

When examining the standardized coefficients of the discriminant functions, the coefficient larger in absolute value indicates the significance of the independent variable. The sign of the coefficient indicates the direction of the relationship. When the table values were examined, it was determined that the independent variables that contributed the most to the separation of classes for Mathematics were " home resources for learning " and "student confident in Math" variables in the first function; "self-efficacy for computer use", "students sense of school belonging" and "student bullying" variables in the second function; "instructional clarity in Math lessons" and "students like learning Math" variables in the third function; and "disorderly behavior during Math lessons" variable in the fourth function.

The standardized coefficients for the discriminant functions obtained for the classification of students for the field of Science were presented in Table 10.

Table 10. Standardized Coefficients for Discriminant Functions (Science)

	Function			
	1	2	3	4
Home_resource	.820*	-.227	-.265	-.339
Confident _S	.474*	-.139	.236	.473
Liking_S	.431	.549*	.086	.085
Self_efficacy_ICT	.347	.355	.562*	.150
School_Belong	.158	.507	-.513*	.282
Ins_Clarity_S	.372	.233	-.221	.746*
Bullying	.252	.282	.094	.368*

When the table values were examined, it was determined that the independent variables that contributed the most to the separation of the classes for the field of Science were "home resources for learning" and "student confident in Science" variables in the first function; "students like learning Science" variable in the second function; "self-efficacy for computer use" and "students sense of school belonging" variables in the third function; and "instructional clarity in Science lessons" and "student bullying" variables in the fourth function.

For the evaluation of the correct classification percentage of the discriminant analysis, the correct classification percentages obtained for Mathematics were given in Table 11.

Table 11. Discriminant Analysis Classification Results (Mathematics)

		International Benchmarks	Predicted Group Membership					Total
			1	2	3	4	5	
Origi	Count	1	196	77	168	11	4	456
		2	113	100	373	86	7	679
		3	66	76	641	264	59	1106
		4	23	27	399	496	139	1084
		5	4	4	71	264	219	562
	%	1	43.0	16.9	36.8	2.4	.9	100.0
		2	16.6	14.7	54.9	12.7	1.0	100.0
		3	6.0	6.9	58.0	23.9	5.3	100.0
		4	2.1	2.5	36.8	45.8	12.8	100.0
		5	.7	.7	12.6	47.0	39.0	100.0

Overall correct classification rate 42.5%.

According to the results obtained from the table, 43% (n=196) of the students assigned to proficiency class 1, 14.7% (n=100) of the students assigned to proficiency class 2, 58% (n=641) of the students assigned to proficiency class 3, 45.8% (n=496) of the students assigned to proficiency class 4 and 39% (n=219) of the students assigned to proficiency class 5 were classified correctly. In total, 1652 students (42.5%) were classified correctly.

For the evaluation of the correct classification percentage of the discriminant analysis, the correct classification percentages obtained for Science were given in Table 12.

Table 12. Discriminant Analysis Classification Results (Science)

		International Benchmarks	Predicted Group Membership					Total
			1	2	3	1	2	
Original	Count	1	167	28	133	19	0	347
		2	96	36	329	109	0	570
		3	64	38	653	486	3	1244
		4	22	18	402	851	31	1324
		5	2	1	53	305	41	402
	%	1	48.1	8.1	38.3	5.5	.0	100.0
		2	16.8	6.3	57.7	19.1	.0	100.0
		3	5.1	3.1	52.5	39.1	.2	100.0
		4	1.7	1.4	30.4	64.3	2.3	100.0
		5	.5	.2	13.2	75.9	10.2	100.0

Overall correct classification rate 45%.

According to the results obtained from the table, 48% (n=167) of the students assigned to proficiency class 1.6% (n=36) of the students assigned to proficiency class 2, 53% (n=653) of the students assigned to proficiency class 3.64% (n=851) of the students assigned to proficiency class 4 and 10% (n=41) of the students assigned to proficiency class 5 were classified correctly. In total, 1748 students (45%) were classified correctly.

3.3. Findings for Comparison Between Logistic Regression and Discriminant Analysis

To make a comparison between logistic regression analysis and discriminant analysis, the correct classification tables obtained with both methods were used. The correct classification tables obtained with both methods were combined in a single table below. The classifications obtained for the comparison of logistic regression and discriminant analysis for the mathematics field were presented in Table 13.

Table 1. Classification Comparison of Logistic Regression and Discriminant Analysis (Mathematics)

	Predicted Group Membership										Correct Classification (%)	
Actual Group Membership	1		2		3		4		5			
	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.
1	138	196	122	77	183	168	12	11	1	4	30,3	43,0
2	86	113	121	100	373	373	95	86	4	7	17,8	14,7
3	45	66	105	76	611	641	310	264	35	59	55,5	58,0
4	9	23	43	27	373	399	546	496	113	139	50,4	45,8
5	1	4	4	4	69	71	286	264	202	219	35,9	39,0

L.R: Logistic regression analysis; D.A: Discriminant analysis

When the table values were examined, it was seen that there were not very serious differences between the results of logistic regression and discriminant analysis. While 138 people were assigned to the correct group as a result of logistic regression analysis for the first proficiency class, 196 people were correctly assigned to this group as a result of discriminant analysis. These values are 121 and 100 for the second proficiency class; 611 and 641 for the third proficiency class; 546 and 496 for the fourth proficiency class and 202 and 219 for the fifth proficiency class; respectively. When the correct classification percentages were examined, the correct classification percentages for the first, third and fifth proficiency classes were high in the discriminant analysis results, while the correct classification percentage results for the second and fourth proficiency classes were high in the logistic regression analysis.

The classifications obtained for the comparison of logistic regression and discriminant analyses for the field of science were presented in Table 14.

Table 2. Classification Comparison of Logistic Regression and Discriminant Analysis (Science)

	Predicted Group Membership										Correct Classification (%)	
Actual Group Membership	1		2		3		4		5			
	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.	L.R.	D. A.
1	99	167	75	28	159	133	14	19	0	0	28,5	48,1
2	50	96	61	36	359	329	100	109	0	0	10,7	6,3
3	29	64	61	38	670	653	476	486	8	3	53,9	52,5

4	8	22	14	18	460	402	800	851	42	31	60,4	64,3
5	1	2	1	1	62	53	293	305	45	41	11,2	10,2

L.R: Logistic regression analysis; D.A: Discriminant analysis

In table 14, it was seen that there were not very serious differences between the results of logistic regression analysis and discriminant analysis, similar to the Mathematics field. For the first proficiency class, 99 people were assigned to the correct group as a result of logistic regression analysis, while 167 people were correctly assigned to this group as a result of discriminant analysis. These values were 61 and 36 respectively for the second proficiency class; 670 and 653 respectively for the third proficiency class; 800 and 851 respectively for the fourth proficiency class; and 45 and 41 respectively for the fifth proficiency class. When the correct classification percentages were examined, while the correct classification percentages for the first and fourth proficiency classes were high in the discriminant analysis results, the correct classification percentage results for the second, third and fifth proficiency classes were high in the logistic regression analysis.

4. Conclusion, Discussion and Suggestions

In this study, ordered logistic regression and discriminant analysis results were examined to determine the variables affecting students' TIMSS 2019 mathematics and science proficiency classifications. First, logistic regression analysis was used in the study and the results were reported. The log likelihood value of the model established by logistic regression analysis in students' 4th grade mathematics and science proficiency classifications showed that model-data fit was achieved. After the validation of the model, independent variables were examined. As a result, for both mathematics and science, students' being bullied at school, having home resources for learning, and perceptions about self-efficacy in computer use were the variables that were effective in achievement classification. In addition, students' liking the course and their self-confidence in the related course were also found to be effective in the achievement classification. Home resources were found to be the most effective variable in the classification made for both mathematics and science. In addition, the highest correct classification rate in Mathematics was obtained for the third proficiency class (55.2%), while the lowest classification rate was obtained for the second proficiency class (17.8%). For Science, the highest correct classification rate was obtained for the fourth proficiency class (60.4%), while the lowest classification rate was observed in the second proficiency class (10.7%).

In the discriminant analysis, all independent variables were found to be significant in predicting the classes. In terms of classification accuracy, the results were in line with logistic regression analysis. When the literature was examined, there were studies in which both analysis methods produce similar results (Abdulqader, 2015; Atar, 2012; Tayyar, 2010). Therefore, both methods can be used in a study to provide evidence of classification validity. However, the assumptions of the discriminant analysis should also be taken into consideration within the scope of the study. If the data do not meet the assumptions, logistic regression analysis can be used, and if the assumptions are met, discriminant analysis can be used if the dependent variable consists of more than two categories.

In mathematics, the variables "sense of school belonging" and "disorderly behavior in Math lessons" among the variables included in the model were not found to be significant; in addition, all other variables included in the model were determined as a significant variable in the classification. In their study, Sarı, Arıkan and Yıldızlı (2017) also examined the school belonging variable among the factors affecting students' Mathematics achievement in the TIMSS 2015 application and as a result of this study, they determined that the school belonging variable had less importance than other variables in predicting students' achievement. On the other hand, only the “instructional clarity” was not significant in science. All other variables considered within the scope of the study were found significant in predicting proficiency classes. It was determined that the variables "bullying", "home resources for learning" and "self-efficacy for internet use" had a significant effect in determining the proficiency classes in both Mathematics and Science. In addition to these, the variables "liking the related course" and "self-confidence in the related course" also had a significant effect on the proficiency classification. Similar findings were observed when the results of the discriminant analysis were analyzed. The most effective variables for both domains are "home resources for learning" and "self-confidence in the related course". These results are consistent with the literature (Aydın-Ceran, 2021; Ayva-Yörü, Sezer-Başaran, & Çakan, 2023; Berger, Holmes & Mackenzie; 2023; Bilican-Demir & Yıldırım, 2021; Büyükgöze & Yakut-Özek, 2023; Geesa, İzci, Song, & Chen, 2019; Karalı et al., 2022).

When the classification performance of logistic regression analysis was examined, 41.6% accuracy rate was obtained for the classification of students into proficiency classes for Mathematics and 43% accuracy rate was obtained for Science. As a result of the discriminant analysis, 42.5% correct classification rate was obtained for mathematics and 45% correct classification rate was obtained for science. Considering the studies on large-scale exams such as TIMSS in the literature, it was expected that the effect of variables on classification will be significant (Akyüz, 2006; Gürsakal, 2012; Karabay, et al., 2015; Özer & Anıl, 2011; Uysal & Yenilmez, 2011; Yavuz, Demirtaşlı, Yalçın, & Dibek, 2017). However, the percentages of the predicted classification percentages were slightly low as a result of the analysis. When analyzed on the basis of proficiency levels, the correct classification rate of the second proficiency level for Mathematics is quite low in both analyses. The third and fourth proficiency levels were correctly classified at a relatively higher rate. In Science, the percentage of correct classification was quite low in the second and fifth proficiency levels in both analyses, while higher percentage of correct classification was obtained in the third and fourth proficiency levels, similar to Mathematics. As a result, the low classification percentages are a point that should be emphasized. This may be due to the limited number of studies with the variables considered in the study. When classification was made according to the independent variables for mathematics, a high proportion of individuals in the first and second proficiency classes were assigned to the third proficiency class. Similarly, in science, a high proportion of individuals in the first and second proficiency class were assigned to the third proficiency class and a large proportion of individuals in the fifth proficiency class were assigned to the fourth proficiency class, resulting in misclassification predictions. Therefore, the variables may have been insufficient to separate individuals.

In this study, independent variables that are thought to affect students' proficiency classifications in Mathematics and Science are discussed. These independent variables are different latent constructs obtained from the measurement tools used in TIMSS 2019. In addition to latent constructs, evidence of classification validity can be obtained with demographic variables or other variables obtained from the TIMSS application. In this way, the percentage of correct classification in both logistic regression analysis and discriminant analysis can be increased. In addition, in this study, math and science data of 4th grade students were used. The analysis can also be conducted using 8th grade data. In this way, the results can be analyzed comparatively.

Kaynakça

Abdulqader, Q. M. (2015). Comparison of Discriminant Analysis and Logistic Regression Analysis: An Application on Caesarean Births and Natural Births Data. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 20* (1), 34-46.

Akyüz, G. (2006). Investigation of The Effect of Teacher and Class Characteristics on Mathematics Achievement in Turkey and European Union Countries. *Elementary Education Online, 5*(2), 75-86.

Atar, H. Y. (2012). Resim-iş öğretmenliği özel yetenek sınavlarının sınıflama doğruluğu üzerine bir çalışma. *Eğitim ve Bilim, 37*(163).

Ceran, S. A. (2021). Evaluation of TIMMS 2019 and PISA 2018 science findings in Turkey perspective. In S.A . Kiray & E. Tomevska-Ilievska (Eds .), *Current Studies in Educational Disciplines 2021* (pp. 78–99). ISRES Publishing

Ayva-Yörü, F. G., Sezer-Başaran, E., & Çakan, M. (2023). Okul, Ev Ve Öğrenci Özelliklerinin 4. Sınıf Öğrencilerinin TIMSS 2019 Matematik Bilişsel Becerilerindeki Rolü. (The Role of School, Home And Student Characteristics in TIMSS 2019 Mathematics Cognitive Skills of 4th Grade Students) *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi,* (67), 197-227. https://doi.org/10.21764/maeuefd.1142768

Berger, N., Holmes, K., & Mackenzie, E. (2023). Instructional Clarity, Classroom Disorder, and Student Achievement in Mathematics: An Exploratory Analysis of TIMSS 2019. In B. Reid-O'Connor, E. Prieto-Rodriguez, K. Holmes, & A. Hughes (Eds.), *Weaving mathematics education research from all perspectives. Proceedings of the 45th annual conference of the Mathematics Education Research Group of Australasia* (pp. 107–114). Newcastle: MERGA.

Bilican-Demir, S., & Yildirim, O. (2021). Indirect effect of economic, social, and cultural status on immigrant students' science performance through science dispositions: A multilevel analysis. *Education and Urban Society, 53*(3), 336-356. https://doi.org/10.1177/0013124520928602

Büyükgöze, H., & Yakut-Özek, B. (2023). The Predictors of Students' Like Mathematics: Confidence, Value, And Instructional Clarity –TIMSS 2019. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, 13*(1), 623-638. https://doi.org/10.30783/nevsosbilen.1230972

Campbell, D. X, & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Coşkun, B., & Karakaya-Özyer, K. (2023). The effect of student characteristics and socioeconomic status on mathematics achievement in Türkiye: Insights from TIMSS 2011-2019*. International Journal of Assessment Tools in Education, 10(*3), 454-481. https://doi.org/10.21449/ijate.1272517

Crocker, L., & Algina, J. (2006). Introduction to classical and modern test theory. Wadsworth.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Cureton, E. E. (1951). Validity.In E. F. Lindquist (Ed.), *Educational measurement* (1st ed., pp. 621-694). Washington, DC:American Council on Education.

Çokluk, Ö., Şekercioğlu G., & Büyüköztürk, Ş., (2012). Sosyal bilimler için çok değişkenli istatistik SPPS ve LISREL uygulamaları. Ankara: Pegem Akademi.

Doğan, N. & Barış, F. (2010). Tutum, Değer Ve Özyeterlik Değişkenlerinin TIMSS-1999 Ve TIMSS-2007 Sınavlarında Öğrencilerin Matematik Başarılarını Yordama Düzeyleri. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(1), 44-50.

Erkuş, A (2004). Geçerlik ve ölçüt dayanak değerlendirme konusunda yeni bir kavramlaştırma önerisi. *Eğitim Araştırmaları Dergisi, 16*, 113-117.

Geesa, R. L., Izci, B., Song, H., & Chen, S. (2019). Exploring Factors of Home Resources and Attitudes towards Mathematics in Mathematics Achievement in South Korea, Turkey, and the United States. *EURASIA Journal of Mathematics, Science and Technology Education, 15*(9), 1-18. https://doi.org/10.29333/ejmste/108487

Güzeller, C. & Kelecioğlu, H. (2006). Ortaöğretim Kurumları Öğrenci Seçme Sınavının Sınıflama Geçerliği Üzerine Bir Çalışma. *H.Ü. Eğitim Fakültesi Dergisi 30*, 140-148.

Karabay, E., Yıldırım, A., & Güler, G. (2015). Yıllara göre PISA matematik okuryazarlığının öğrenci ve okul özellikleri ile ilişkisinin aşamalı doğrusal modeller ile analizi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 1*(36), 137-151.

Karali, Y., Varol-Palancioglu, Ö., & Aydemir, H. (2022). Examining the Factors Affecting Turkey's 4th Grade Mathematics Achievement According to TIMSS 2019 Final Report. *International Journal of Curriculum and Instruction, 14*(1), 424-454.

Karasar, N. (2014). Bilimsel Araştırma Yöntemi. Ankara: Nobel.

Kan, A. (2004). ÖSS'nin Sınıflama Geçerliği Üzerine Bir Çalışma. *İnönü Üniversitesi Eğitim Fakültesi Dergisi,5*(8).

Kane, M.T. (2001). Current concerns in validity theory, *Journal of Educational Measurement, 38*(4), 319-342.

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology, 101*(1), 190-206. https://doi.org/10.1037/a0013153

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Murphy, R. K. ve Davidshofer, O. C. (2005). Psychological Testing: Principles and Application Sixth Edition. New Jersey: Prentice-Hall.

Özer, Y., & Anıl, D. (2011). Öğrencilerin fen ve matematik başarilarini etkileyen faktörlerin yapisal eşitlik modeli ile incelenmesi*. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 41*(41), 313-324.

Saral, D. G. (2012). A study upon classification validity of decisions taken with regard to norm and criterion. Master's thesis, Hacettepe University: Ankara

Sarı, M. H., Arıkan, S. ve Yıldızlı, H. (2017). Factors predicting mathematics achievement of 8th graders in TIMSS 2015*. Journal of Measurement and Evaluation in Education and Psychology, 8*(3), 246-265. https://doi.org/10.21031/epod.303689

Schmidt, F. L. (2012). Cognitive tests used in selection can have content validity as well as criterion validity: A broader research review and implications for practice. *International Journal of Selection and Assessment, 20*(1), 1-13. https://doi.org/10.1111/j.1468-2389.2012.00573.x

Şahin, M. G., Çelik, Ö. C., & Yıldırım, Y. (2022). The relationships between teacher and student qualities in TIMSS 2019: A path analysis model*. Psycho-Educational Research Reviews 11*(3), 443-461. https://doi.org/10.52963/perr_biruni_v11.n3.04

Şevgin, H., & Eranıl, A. K. (2023). Investigation of Turkish Students' School Engagement through Random Forest Methods Applied to TIMSS 2019: A Problem of School Psychology. *International Journal of Psychology and Educational Studies, 10*(4), 896-909. https://doi.org/10.52380/ijpes.2023.10.4.1260

Tabachnick, B. G., & Fidell, L.S. (2013). Multivariate statistics. New Jersey: Pearson Education.

Taşdemir, F. (2015). A Study for developing a success test: examination of validity and classification accuracy by ROC analysis. *Procedia-Social and Behavioral Sciences, 191*, 110-114. https://doi.org/10.1016/j.sbspro.2015.04.473

Tayyar, N. (2010). Predicting Customer Satisfaction by Comparing Performances of Artificial Neural Networks, Logistic Regression and Discriminant Analysis. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15*(1), 339-355.

Tekin H. (1977). Eğitimde Ölçme ve Değerlendirme. Ankara: Mars Matbaası.

Upadyaya, K., & Salmela-Aro, K. (2013). Development of school engagement in association with academic success and well-being in varying social contexts. *European Psychologist, 18*(2), 136-147. https://doi.org/10.1027/1016-9040/a000143

Uysal, E., & Yenilmez, K. (2011). Sekizinci sınıf öğrencilerinin matematik okuryazarlığı düzeyi. (The Mathematics Literacy Level of Eighth Grade Students). *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 12*(2), 1-15.

Yavuz, H. Ç., Demirtaşlı, R. N., Yalçın, S. & İlgün Dibek, M. (2017). The Effects of Student and Teacher Level Variables on TIMSS 2007 and 2011 Mathematics Achievement of Turkish Students. *Eğitim ve Bilim, 42*(189), 27-47. http://dx.doi.org/10.15390/EB.2017.6885

Yin, L., & Fishbein, B. (2019). Creating and interpreting the TIMSS 2019 context questionnaire scales. *Methods and procedures: TIMSS, 16*(1), 1-331.