



(ISSN: 2602-4047)

Sirganci, G. (2023). Comparison of GPCMLasso and Alignment Methods in Detecting Differential Item Functioning, *International Journal of Eurasian Education and Culture*, 8(23), 2270-2299.

DOI: <http://dx.doi.org/10.35826/ijoec.764>

Article Type (Makale Türü): Research Article

COMPARISON OF GPCMLASSO AND ALIGNMENT METHODS IN DETECTING DIFFERENTIAL ITEM FUNCTIONING

Gözde SIRGANCI

Dr., Southern Methodist University, Dallas, US, gsirganci@smu.edu

ORCID: 0000-0003-4824-5413

Received: 21.04.2023

Accepted: 18.09.2023

Published: 01.10.2023

ABSTRACT

The objective of this study is to evaluate the effectiveness of the proposed GPCMLasso and Alignment methodologies in identifying Differential Item Functioning (DIF), by comparing their performances using real-world data. Specifically, the study focuses on the Interest in ICT subscale from the 2018 Program for International Student Assessment (PISA) ICT questionnaire. The analysis includes data from five Eastern European and Central Asian countries that participated in the PISA 2018 ICT questionnaire: Turkey, Bulgaria, Croatia, Georgia, and Kazakhstan, covering a sample of 29,277 15-year-old students. According to the Alignment method, the factor loadings for items 2, 3, 4, and 6, as well as the factor intercepts for items 2 and 6, remain consistent across all countries. In contrast, the GPCMLasso approach indicates that every item on the measurement tool exhibits both DIF and Differential Step Functioning (DSF) across various countries. The study finds that there is a 64% agreement rate between the two methods in detecting DIF. However, the GPCMLasso method appears to be more sensitive in identifying Differential Metric Functioning (DMF) compared to the Alignment method.

Keywords: Differential item functioning, differential step functioning, regularization, machine learning.

INTRODUCTION

Validity refers to how accurately and comprehensively a particular trait is measured (Mellor, 1995). Sometimes, variables like language, gender, ethnicity, culture, socio-economic status, or even specific cognitive abilities like problem-solving can affect the latent trait being assessed. This kind of interference compromises the validity of the test and leads to biased scores (Penfield & Camilli, 2006). To examine such bias, Differential Item Functioning (DIF) analyses are employed. DIF explores the varying likelihood of individuals from different backgrounds but with the same ability level answering a test item correctly (Penfield & Camilli, 2006). Various methods, both traditional (Chang et al., 1996; Choi et al., 2011; Holland & Thayer, 1988; Swaminathan ve Rogers, 1990; Thissen ve et al., 1993) and new approaches (Belzak, 2023; Cohen & Bolt, 2005; De Ayala et al., 2002; Muthén & Asparouhov, 2010; Schauburger & Mair, 2020), have been developed to detect DIF. Traditional approaches like Mantel-Hanzel and SIBTEST often determine DIF based on a single variable that is either binary or multi-categorical and cannot conduct DIF analyses for continuous variables such as age and response time. Furthermore, when analyses are conducted based on a single variable, the situation where multiple background characteristics of participants might together be a source of DIF is overlooked (Belzak, 2023). In addition, these techniques use exact zero constraints (i.e., exact equality of parameters between groups) when comparing parameters between groups (Muthén & Asparouhov, 2012). Particularly in this case, especially in multi-group comparisons ($n \geq 10$), if the relevant item shows strong DIF between a limited number of groups, it can appear biased for all groups (Van de Vijver et al., 2019). Another issue is that multiple potential sources of DIF can affect test responses and scores in various linear, non-linear, and even non-parametric ways, and many of these cannot be detected with these methods (Belzak, 2023; Schauburger & Mair, 2020). To overcome the aforementioned limitations of DIF detection with traditional methods, various regularization or bias correction methods known under approximate measurement invariance methods as structural equation modeling-based methods (Muthén & Asparouhov, 2012; Asparouhov & Muthén, 2014) and Machine Learning (ML) approaches have been suggested in the literature (Belzak & Bauer, 2020; Liang & Jacobucci, 2020; Tutz & Schauburger, 2015; Schauburger & Tutz, 2016; Strobl et al., 2015). While approximate measurement invariance methods focus especially on determining the most appropriate measurement model in large-scale assessments where multi-group comparisons are made (Muthén & Asparouhov, 2012); ML methods have focused on evaluating multiple variables together in linear, non-linear, and non-parametric formats of DIF (Belzak, 2023).

In conclusion, DIF is a significant validity issue (Xu & Tracey, 2017), and in practices where high-risk decisions are made based on test scores, the validity of these scores needs to be established. For this purpose, it is essential to comparatively examine the many existing techniques and identify their strengths and weaknesses. Unlike traditional DIF identification methods, new approaches are valuable both in terms of relaxing the strict assumptions required by traditional methods and in identifying the primary source of DIF by discovering linear or non-linear relationships between multiple potential sources of DIF. In this context, the problem of this study is to compare the alignment approach based on structural equation modeling and the machine learning-based lasso penalty approach to detect DIF and determine the strengths and weaknesses of these methods.

METHOD

Research Model

Descriptive studies examine individuals, settings, conditions, or events from a current point of view and interpret the data they provide. It focuses solely on the phenomenon of interest as it exists naturally without manipulating individuals, conditions, or events (Mertler, 2014). A descriptive research method was employed in this study, in which alignment and GPCMLasso DIF methods were compared using the PISA 2018 *Interest in ICT* questionnaire.

Data Source

The OECD (2021) report emphasizes that the lack of education and learning opportunities in Eastern European and Central Asian (EECA) countries threatens the future of children and that in order to reform education, EECA countries need to understand the performance of their own education systems and compare their results with those of other countries. In this context, analyses were carried out with participants from Eastern European and Central Asian (EECA) countries. The PISA-EECA countries have continuously increased their involvement, and ten countries participated in 2018: Baku (Azerbaijan), Belarus, Bulgaria, Croatia, Georgia, Kazakhstan, Moldova, Romania, Turkey, and Ukraine. However, five out of 10 countries participated in the ICT questionnaire (OECD, 2021). A summary of EECA countries' information is provided in Table 1.

Table 1. Frequencies and Percentages Related to the Study Group

	Country	Gender	f	%	N
BRG	Bulgaria	Female	1276	49.6	2574
		Male	1298	50.4	
GEO	Georgia	Female	1336	51.1	2613
		Male	1277	48.9	
HRV	Croatia	Female	2508	51.9	4835
		Male	2327	48.1	
KAZ	Kazakhstan	Female	6724	48.8	13775
		Male	7051	51.2	
TUR	Turkey	Female	2736	49.9	5480
		Male	2744	50.1	

Data Collection Tool

In this study, the data from the PISA 2018 Interest in ICT questionnaire were used (OECD, 2019a). A four-point Likert rating scale was utilized (1 = "strongly disagree", 2 = "disagree", 3 = "agree", 4 = "strongly agree"). The descriptive statistics and reliabilities and CFA model fit indexes of the Interests in ICT questionnaire for all countries are illustrated in Table 2.

Table 2. Mean, Standard Deviation, Reliabilities and CFA model fit index of the Interests in ICT questionnaire

Country	(mean±sd)	α	CFI	TLI	SRMR
BRG	4.33±1.1	.90	.96	.97	.050
GEO	4.48±1.1	.90	.98	.97	.022
HRV	3.34±0.8	.82	.97	.95	.038
KAZ	3.57±0.9	.85	.96	.94	.035
TUR	4.29±1.1	.88	.97	.98	.045

Data Analysis

Alignment

The alignment approach assesses measurement invariance by generating a model that permits limited variations in measurement parameters across different groups. Rather than assuming invariance exists, the method seeks the most fitting pattern of measurement invariance and then calculates factor means and variances for each group. One key benefit of this approach is that it is rooted in structural modeling, enabling the prediction of models suitable for a large array of groups without sacrificing model fit. The alignment technique estimates factors like loadings (λ_g), intercepts (v_g), means (α_g), and variances (ψ_g) by identifying the number of fluctuating item parameters and the model that minimizes deterioration in measurement invariance (Asparouhov & Muthén, 2014).

The testing process in the alignment method occurs in two phases. Initially, a foundational model, labeled "M₀," is created where factor loadings and cutoffs are group-specific, while factor averages are set at zero and variances at one for all groups. This initial model is considered optimal since it does not enforce any inter-group parameter constraints. The next phase involves alignment optimization, where factor means and variances that minimize distortion in measurement invariance are computed using a simplicity function; equation 1 (Asparouhov & Muthén, 2014).

$$F = \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{pg_1} - \lambda_{pg_2}) + \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(v_{pg_1} - v_{pg_2}) \tag{1}$$

$$w_{g_1, g_2} = \sqrt{N_{g_1} N_{g_2}} \tag{2}$$

w: factor weights, N: sample size

The finalized aligned model, known as "M₁," retains the same fit quality as its foundational counterpart, M₀. Even though the model strives to minimize the level of invariance, it does not compromise the model's overall fit. Asparouhov and Muthén(2014) suggest that to deem the model's fit as acceptable, no more than 25% of the items should violate statistical invariance; Flake and McCoach (2018) set this limit at 29%. Another metric used to evaluate model fit is the R² effect size measure, which reveals the extent to which variability in item parameter estimation is accounted for by the group-specific factor means and variances. An R² value close to 1 signifies

complete invariance, while an R^2 value around 0 means group differences do not explain the item parameter variability (Asparouhov & Muthén, 2014). This study employed Mplus 8.0 software for conducting DIF analyses using the alignment method.

GMPClasso

The Generalized Partial Credit Model with Lasso regularization (GPCMlasso) forms a unified model that uses penalized likelihood estimation for parameter selection in detecting Differential Item Functioning (DIF), as described by Schauburger & Mair (2020). One of the major advantages of this approach is its ability to consider multiple variables simultaneously and its applicability to both continuous and categorical variables for DIF detection. With the GPCMlasso model, uniform DIF is calculated according to country (G) variable, as indicated in Equation 3.

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) = \beta_i = [\theta_p + x_p^T \alpha - \delta_{ir} - (\gamma_i \times G)] \tag{3}$$

In this model γ_i represents DIF parameters, are the effects of grouping variable (G) on item i , respectively. If these parameters are not equal to zero after applying lasso penalization, they are considered uniform DIF. For the parameter estimate, the GPCMlasso model figure out the following lasso penalized log-likelihood function illustarted equation 4.

$$\ell_p(\theta, \alpha, \delta, \beta, \gamma) = \ell(\theta, \alpha, \delta, \beta, \gamma) - \lambda \sum_{i=1}^l \sum_{j=1}^m w_{ij} |\gamma_{ij}| \tag{4}$$

where $\ell(\theta, \alpha, \delta, \beta, \gamma)$ is the regular version of the log-likelihood function, $\lambda \sum_{i=1}^l \sum_{j=1}^m w_{ij} |\gamma_{ij}|$ represents the lasso penalty term and $\lambda \geq 0$ is the tuning parameter that controls the degree of penalization applied to the vector of regression coefficients γ_{ij} . In this particular model, x_p denotes an m dimensional vector of covariates for individual p (G), and θ stands for the latent construct in focus. Furthermore, the DIF parameter, γ_i , the item step parameter δ_{ij} , the main effect α and, the item discrimination β_i parameters are estimated for item i . Penfield (2007) and Penfield et al. (2009) pointed out, that DIF analysis in polytomous settings can also suffer from the phenomenon of differential step functioning (DSF). DSF implies differences at the level of each step of a polytomous item with respect to different covariate values. Therefore, in this study, DAF was also examined together with DMF. When using penalized likelihood approaches like this one, determining the optimal tuning parameter λ is a key step. To identify the best value for λ , it can either be used model selection criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), or cross-validation (CV)

techniques. In this study, emphasis was primarily placed on BIC, since selecting the optimal variables is more critical than making accurate predictions when it comes to assessing DIF. The BIC is specified for the GPCMLasso model in Equation 5.

$$BIC(\lambda) = -2L\lambda(.) + df(\lambda)\log(n) \tag{5}$$

The GPCMLasso package in R software (R Core Team, 2018) was used to assess DIF.

FINDINGS

The findings are presented in two stages, first the alignment findings, and then the GPCMLasso findings.

Alignment Findings

Before conducting alignment analyses, a Multiple Group Confirmatory Factor Analysis (MG-CFA) was performed. Table 3 provides the model fit statistics for the structural, metric, and scalar invariance levels that emerged from the MG-CFA. Chi-square difference tests reveal that there is a significant decline in model fit when moving from structural to metric invariance, as well as from metric to scalar invariance. When looking at the shifts in alternate goodness-of-fit indices such as CFI, RMSEA, and SRMR, the differences surpass acceptable levels both in the transition from structural to metric invariance and from metric to scalar invariance. These results indicate that the average factor scores for the six-item interest in ICT interest questionnaire are not comparable among the five countries being studied.

Table 3. MG-CFA Results

Model	χ^2	df	p	RMSEA	CFI	SRMR
Yapısal	269.557	45	<.001	.095 [.085-.106]	.972	.031
Zayıf	390.661	65	<.001	.096 [.087-.105]	.960	.080
Güçlü	838.929	85	<.001	.127 [.119-.135]	.907	.078

In Table 4, the data about the invariance of threshold and factor loadings for each question in the 6-item interest in ICT questionnaire across countries is shown. Parenthetical notations indicate significant differences in factor loadings and thresholds for the respective country. The factor loadings for items 2, 3, 4, and 6 are invariant across countries. However, for item 1, there are significant differences in the factor loadings for Georgia and Kazakhstan. Annex 1 provides detailed results of country-by-country comparisons for items where invariance is not achieved, both in terms of factor loadings and intercept. Specifically, Georgia's factor loading for this item is significantly different from those of Turkey and Croatia, while Kazakhstan's factor loading differs significantly from all countries in the study. For item 5, only Kazakhstan shows a significant difference in factor loading. According to the pairwise comparison results, this difference is with Turkey and Bulgaria. In general, only 3% of the factor loadings for interest in ICT items diverge significantly across countries, which is within the acceptable range of less than or equal to 25% as suggested by Asparouhov & Muthén (2014). As a result, factor loadings can be

compared between countries, confirming metric invariance for the interest in ICT questionnaire. When looking at the factor intercepts for the interest in ICT items across countries, it appears that the intercepts for item 2 and item 4 are consistent across all countries. However, variations are observed for other items. Specifically, for item 1, the intercepts differ significantly for Turkey and Kazakhstan. For item 3, it differs for Turkey. For item 4, Croatia, Georgia, and Kazakhstan show differences. Finally, for item 5, measurement invariance does not hold in Croatia and Kazakhstan. Overall, when examining the intercepts for ICT-related items among countries, 27% of these intercepts deviate significantly between groups, exceeding the 25% guideline from Asparouhov & Muthén (2014). Hence, the interest in ICT questionnaire does not hold scalar invariance, making it inappropriate to compare average factor scores between countries based on the overall scale score.

Table 4. Alignment Results for Factor Loadings and Intercepts

Item	Intercepts	Count of Non-invariant parameter	Factor Loadings	Count of Non-invariant parameter
13Q01 (<i>item1</i>)	1 (2) 3 4 (5)	2	1 2 3 (4) (5)	2
13Q04 (<i>item2</i>)	1 2 3 4 5	0	1 2 3 4 5	0
13Q05 (<i>item3</i>)	1 (2) 3 4 5	1	1 2 3 4 5	0
13Q11 (<i>item4</i>)	1 2 (3) (4) (5)	3	1 2 3 4 5	0
13Q12 (<i>item5</i>)	1 2 (3) 4 (5)	2	1 2 3 4 (5)	1
13Q13 (<i>item6</i>)	1 2 3 4 5	0	1 2 3 4 5	0
Toplam		8(%26.6)		3 (%10)

*BGR=1, TUR=2, HRV=3, GEO=4, KAZ=5

In Table 5, the factor means estimated by the alignment method are presented. Accordingly, it is observed that Croatia has the highest factor mean, while Kazakhstan has the lowest. The factor means between Croatia and other countries are significantly different from each other. Similarly, the factor means of Kazakhstan and Turkey are significantly higher than those of Bulgaria and Georgia.

Table 5. Comparison of Factor Means Between Countries

Ranking	Group	Factor Mean	Groups With Significantly Smaller Factor Mean
1	3	0.431	5 2 1 4
2	5	0.169	1 4
3	2	0.130	1 4
4	1	0.000	
5	4	-0.078	

*BGR=1, TUR=2, HRV=3, GEO=4, KAZ=5

GPCMlasso Findings

Table 6 represents the results of the GPCMlasso model for detecting DIF for country variables based on BIC criteria. The variable country is encoded by dummy coding with Bulgaria as the reference category. Therefore, all the results related to the country variable in Table 6 reflect the results of the pairwise comparison of each country with Bulgaria.

Table 6. DIF Results for Country, Gender, and ESCS Variables Based on GPCMLasso Coefficients

	Code	CNT_GEO	CNT_HRV	CNT_KAZ	CNT_TUR
Interest in ICT	13Q01 (item1)	0	-0.029	0.204	-0.195
	13Q04 (item2)	0	0	-0.035	0
	13Q05 (item3)	-0.02	0	0	0.09
	13Q11 (item4)	-0.084	0.111	-0.117	0
	13Q12 (item5)	0	0.114	0.147	0.042
	13Q13 (item6)	0	-0.016	0	0

In the context of lasso regression, the sign of the coefficient represents the direction of the relationship between the predictor and the outcome variable. A positive coefficient indicates that the item is biased in favor of the focal group, while a negative coefficient indicates that the item is biased against the focal group. It is important to remember that the coefficient's magnitude also provides information about the strength or size of the DIF effect. The bigger the absolute value of the coefficient, the stronger the DIF effect is. Thus, both the sign and magnitude of the lasso coefficient are essential for understanding and interpreting DIF in the context of the examined items and groups. In Figure 1, graphs are presented that show the parameter paths followed by DMF parameters for each item along the λ tuning parameter when the penalty term is applied. Paths are drawn separately for each item. The red dashed lines represent the model that is most suitable according to BIC. In this way, if the parameter path of a specific variable crosses the vertical red dashed line, the relevant item contains DMF. Findings related to the DIF according to the country variable for the relevant item are presented on the left side of Figure 1, while findings related to the DAF are presented on the right side. Table 6 and Figure 1 have been evaluated together while presenting the GPCMLasso findings.

All items of the interest in BIT questionnaire contain DIF according to the country variable. Item 1 has DIF in favor of Kazakhstan between Bulgaria and Kazakhstan; in favor of Bulgaria between Bulgaria and Croatia; and in favor of Bulgaria between Bulgaria and Turkey. When examining the DSF, the second and third step parameters of item 1 have different functions in favor of Kazakhstan between Bulgaria and Kazakhstan; and in favor of Bulgaria between Bulgaria and Turkey. On the other hand, although item 1 shows DIF between Bulgaria and Croatia, it does not show DSF. The reason for this could be that the DIF is negligible. Item 2 only shows DIF in favor of Bulgaria between Bulgaria and Kazakhstan. On the other hand, according to DSF results, the first step parameter of item 2 shows DSF in favor of Kazakhstan, and the second step parameter shows DAF in favor of Bulgaria. Also, despite there being no DIF between Bulgaria and Kazakhstan, the second step parameter of the item shows DSF in favor of Bulgaria between these two countries. Item 3 contains both DIF and DSF for all step parameters in favor of Turkey between Bulgaria and Turkey. On the other hand, Item 3 shows negligible DIF in favor of Bulgaria between Bulgaria and Georgia and does not show DSF. Item 4 shows DIF in favor of Croatia between Bulgaria and Croatia and shows DAF for the second and third step parameters. Item 4 also contains DIF in favor of Bulgaria for both Georgia and Kazakhstan. There is DSF for the second step function for both countries. Item 5 contains DIF in favor of Kazakhstan between Bulgaria and Kazakhstan; and DSF for the second and third step parameters. It also contains DIF and DSF for all step parameters in favor of Croatia between Bulgaria and Croatia. It contains

DIF in favor of Turkey and DSF for the first and second step parameters between Bulgaria and Turkey. Item 6 contains DIF in favor of Bulgaria and DSF for the second and third step parameters between Bulgaria and Croatia.

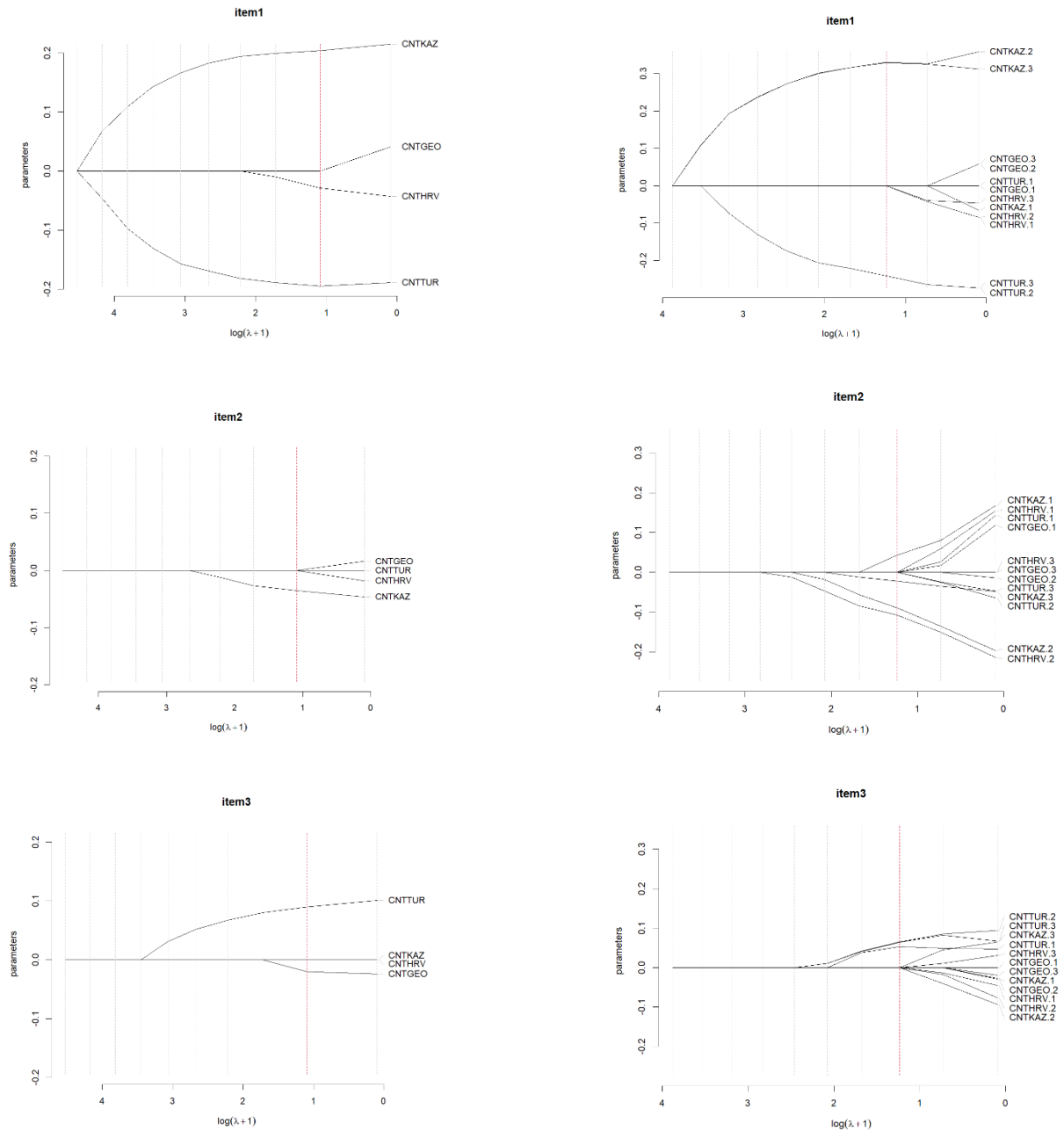


Figure 1. Coefficient paths of all DIF and DSF parameters for GPCmlasso for interest in ICT

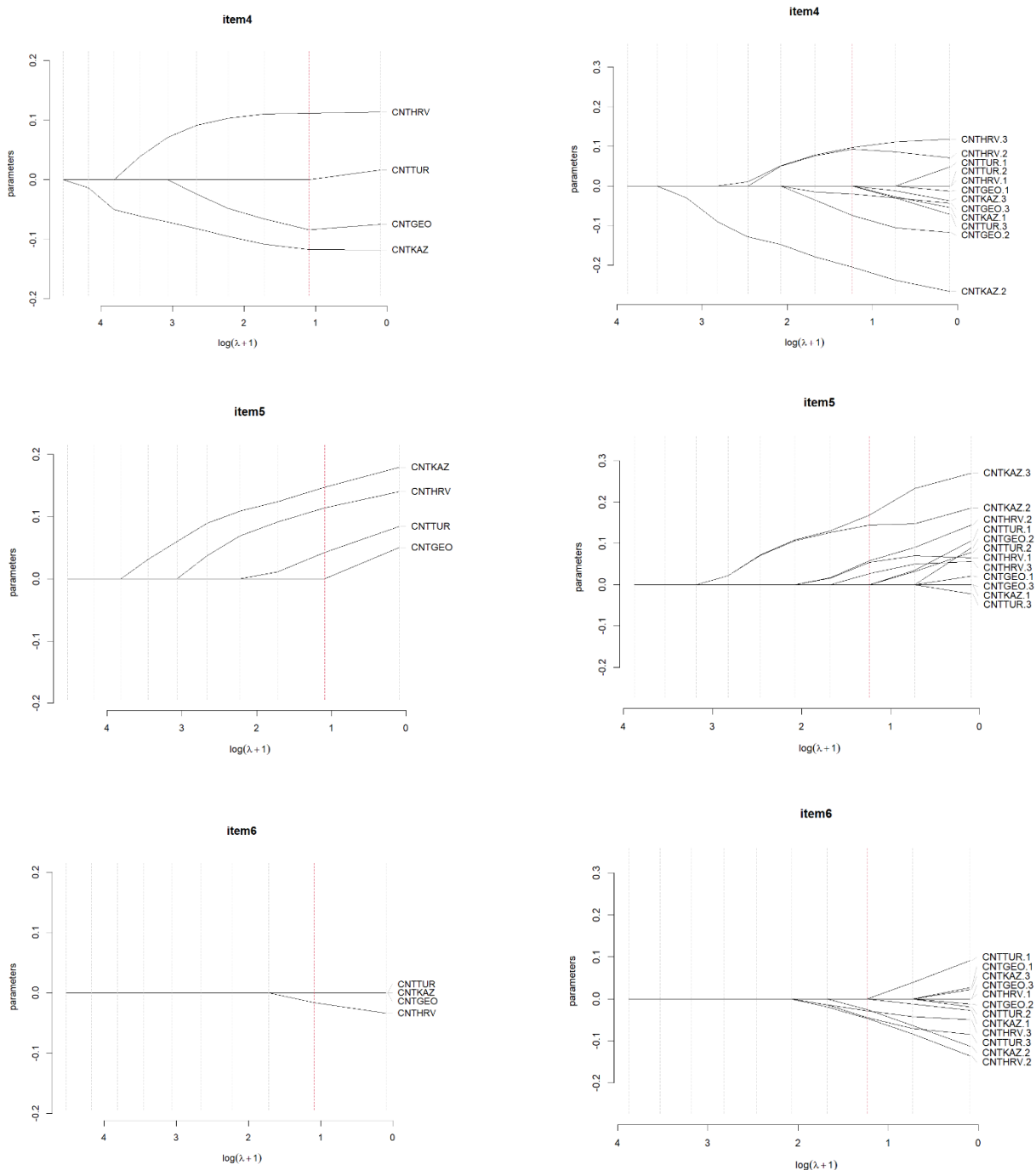


Figure 1. Coefficient paths of all DIF and DSF parameters for GPCMLasso for interest in ICT (Continue)

In Table 7, DIF results are presented, comparing the reference country, Bulgaria, with other countries. These results are derived from both the Alignment method and the GPCMLasso approach. A 64% agreement rate was found between the two methods in identifying DIF. According to the Alignment method, items 2 and 6 show no DIF among any of the countries compared. In contrast, the GPCMLasso method indicates that item 2 has DIF between Bulgaria and Kazakhstan, and item 6 has DIF between Bulgaria and Croatia. Despite these differences,

the magnitude of DIF in these specific items is considered negligible. Therefore, it can be concluded that GPCMLasso is a more sensitive tool for detecting DIF compared to the Alignment method.

Table 7. Summary of DIF Results based on GPCMLasso and Alignment

	Item 1		Item 2		Item 3		Item 4		Item 5		Item 6	
	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI
BGR-TUR	+	+			+	+			+	+		
BGR-HRV		+					+	+	+	+		(+)
BGR-GR						+	+	+	+			
BGR-KAZ	+	+		(+)			+	+	+	+		

CONCLUSION and DISCUSSION

In this study, the alignment and GPCMLasso methods are compared to detect DIF of the PISA 2018 interest in ICT questionnaire among Eastern European and Central Asian countries. While the alignment method is based on structural equation modeling, GPCMLasso is based on a machine-learning approach. While factor loadings of items 2, 3, 4, and 6 hold invariance, factor intercept of items 2 and 6 hold invariance across all countries. It was concluded that interest in ICT scale did not hold scalar invariance, since the rate of parameters that distort invariance was 3% among factor loads and 27% ($\geq 25\%$, Asparouhov & Muthén, 2014) among factor intercepts. According to this result, the factor mean of the interest in ICT cannot be compared among countries. There are a few studies in the literature examining the measurement invariance of the PISA-ICT questionnaire across countries. Odell et al. (2021) used the alignment technique to examine the measurement invariance of the PISA 2015 ICT survey on 47 countries and stated that it is not possible to compare the mean scores obtained from the ICT among countries. This finding is consistent with the research findings obtained from this study. On the other hand, Ma and Qin (2021) tested the measurement invariance of the PISA 2018 ICT survey across 16 countries with multi-group confirmatory factor analysis and found that residual (strict) invariance was achieved.

The GPCMLasso results showed that all items of the PISA 2018 Interest in ICT questionnaire included DIF and DSF across countries. Two items (3,4) between the reference group Bulgaria and Georgia, four items between Bulgaria and Croatia, (1,4,5,6); four items (1,2,4,6) between Bulgaria and Kazakhstan and three items (1, 3,5) between Bulgaria and Turkey contain DIF. In addition, GPCMLasso provides information about the significant differences in category parameters of polytomous items among groups by performing differential step functioning. In this study, it has been observed that there are different functionalizations, especially in the second and third-category parameters of the reference country Bulgaria and the countries showing DIF.

In conclusion, both methods have similarly identified items that contain DIF according to the country variable in the PISA 2018 Interest in CIT questionnaire. The GPCMLasso, on the other hand, is more strict in detecting DIF due to the fact that it captures a negligible amount of DIF. While the alignment approach presents pairwise comparison results between all countries in a single analysis, the GPCM lasso method presents pairwise comparison results according to the reference country. On the other hand, the GPCM lasso method stands out

in terms of calculating which item response category level there is group differentiation among polytomous items.

SUGGESTIONS

In this research, DIF analyses have been carried out on the PISA 2018 Interest in BIT questionnaire, which is unidimensional using alignment and GPCMLasso methods. However, both methods can perform simultaneous analysis in multidimensional structures. Therefore, future studies can investigate the performance of these methods in multidimensional frameworks. Additionally, this study has compared new approaches in DIF detection, such as the structural equation modeling-based alignment approach and the machine learning-based GPCMLasso method. Comparisons between methods could also include IRT-based DIF detection approaches. Hence, a simulation study that examines methods based on different approaches may be beneficial to understanding which method can most accurately identify DIF under specific conditions.

ETHICAL TEXT

This article followed the journal writing rules, publication principles, research and publication ethics, and ethical journal rules. The author is responsible for any violations that may arise regarding the article. An ethics committee is not required for this study.

Conflict of interest: The author declared no conflict of interest.

Financial Support: There is no financial support from any institution or organization in this study.

Author Contribution Rate: In this study, the contribution rate of the sole author was 100%.

REFERENCES

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508. <https://doi.org/10.1080/10705511.2014.919210>
- Belzak, W. C. M., & Bauer, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychological Methods*, 25(6), 673–690. <https://doi.org/10.1037/met0000253>
- Belzak, W. C. (2023). The Multidimensionality of Measurement Bias in High-Stakes Testing: Using Machine Learning to Evaluate Complex Sources of Differential Item Functioning. *Educational Measurement: Issues and Practice*, 42(1), 24-33. <https://doi.org/10.1111/emip.12486>
- Chang, H. H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items an adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333–353. <https://doi.org/10.1111/j.1745-3984.1996.tb00496.x>

- Choi, S., Gibbons, L., & Crane, P. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(1), 1–30. <https://doi.org/10.18637/jss.v039.i08>
- Flake, J. K., & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1), 56-70. <https://doi.org/10.1080/10705511.2017.1374187>
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Lawrence Erlbaum Associates, Inc.
- Liang, X., & Jacobucci, R. (2020). Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 722-734. <https://doi.org/10.1080/10705511.2019.1693273>
- Ma, Y., & Qin, X. (2021). Measurement invariance of information, communication and technology (ICT) engagement and its relationship with student academic literacy: Evidence from PISA 2018. *Studies in Educational Evaluation*, 68, 100982. <https://doi.org/10.1016/j.stueduc.2021.100982>
- Mellor, L. T. (1995). *A comparison of four differential item functioning (DIF) methods for polytomously scored items* (Order No. 9534889). Available from ProQuest One Academic. (304235582). <http://proxy.libraries.smu.edu/login?url=https://www.proquest.com/dissertations-theses/comparison-four-differential-item-functioning-dif/docview/304235582/se->
- Mertler, C. A. (2014). *The Data-Driven Classroom: How do I use student data to improve my instruction?* (ASCD Arias). ASCD.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <https://doi.org/10.1037/a0026802>
- Odell, B., Gierl, M., & Cutumisu, M. (2021). Testing measurement invariance of PISA 2015 mathematics, science, and ICT scales using the alignment method. *Studies in Educational Evaluation*, 68, 100965. <https://doi.org/10.1016/j.stueduc.2020.100965>
- OECD, U. (2021). *Education in Eastern Europe and Central Asia: Findings from PISA*.
- Penfield, R. D., & Camilli, G. (2006). Differential Item Functioning and Item Bias. *Handbook of statistics*, 26, 125-167. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X)
- Penfield, R. D. (2007). Assessing differential step functioning in polytomous items using a common odds ratio estimator. *Journal of Educational Measurement*, 44(3), 187–210. <https://doi.org/10.1111/j.1745-3984.2007.00034.x>
- Penfield, R. D., Gattamorta, K., & Childs, R. A. (2009). An NCME instructional module on using differential step functioning to refine the analysis of DIF in polytomous items. *Educational Measurement: Issues and Practice*, 28(1), 38–49. <https://doi.org/10.1111/j.1745-3992.2009.01135.x>
- Schauberger, G., Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behav Res* 52, 279–294 (2020). <https://doi.org/10.3758/s13428-019-01224-2>
-

- Schauberg, G., & Tutz, G. (2016). Detection of differential item functioning in Rasch models by boosting techniques. *British Journal of Mathematical and Statistical Psychology*, 69(1), 80–103. <https://doi.org/10.1111/bmsp.12060>
- Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289–316. <https://doi.org/10.1007/s11336-013-9388-3>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In Holland, P. W., & Wainer, H. (Eds.) *Differential item functioning*, (pp. 67-113).
- Tutz, G., & Schauberg, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika*, 80(1), 21–43. <https://doi.org/10.1007/s11336-013-9377-6>
- Xu, H., & Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *The European Journal of Counselling Psychology*, 6(1), 75–82. <https://doi.org/10.5964/ejcop.v6i1.120>
- Van de Vijver, F. J., Avvisati, F., Davidov, E., Eid, M., Fox, J. P., Le Donné, N., ... & van de Schoot, R. (2019). Invariance analyses in large-scale studies. <https://doi.org/10.1787/19939019>

APPENDIX

Appendix 1. Pairwise comparison all countries

	Group	Group	Intercept					Loadings				
			Value	Value	Difference	SE	Sig.	Value	Value	Difference	SE	Sig.
Item1	2	1	2.677	2.418	0.258	0.046	0	0.742	0.624	0.118	0.048	0.015
	3	1	2.424	2.418	0.006	0.039	0.879	0.707	0.624	0.083	0.061	0.176
	3	2	2.424	2.677	-0.253	0.048	0	0.707	0.742	-0.035	0.058	0.551
	4	1	2.349	2.418	-0.069	0.046	0.136	0.545	0.624	-0.079	0.045	0.08
	4	2	2.349	2.677	-0.327	0.049	0	0.545	0.742	-0.197	0.047	0
	4	3	2.349	2.424	-0.075	0.047	0.115	0.545	0.707	-0.162	0.061	0.008
	5	1	2.135	2.418	-0.283	0.043	0	0.338	0.624	-0.286	0.051	0
	5	2	2.135	2.677	-0.542	0.048	0	0.338	0.742	-0.404	0.054	0
	5	3	2.135	2.424	-0.289	0.044	0	0.338	0.707	-0.369	0.066	0
	5	4	2.135	2.349	-0.215	0.045	0	0.338	0.545	-0.207	0.051	0
Item3	2	1	2.656	2.8	-0.144	0.036	0	0.631	0.721	-0.09	0.034	0.008
	3	1	2.76	2.8	-0.039	0.034	0.248	0.702	0.721	-0.019	0.04	0.632
	3	2	2.76	2.656	0.105	0.041	0.011	0.702	0.631	0.071	0.048	0.14
	4	1	2.826	2.8	0.027	0.029	0.352	0.685	0.721	-0.036	0.026	0.169
	4	2	2.826	2.656	0.171	0.037	0	0.685	0.631	0.054	0.035	0.127
	4	3	2.826	2.76	0.066	0.035	0.056	0.685	0.702	-0.017	0.04	0.671
	5	1	2.775	2.8	-0.025	0.027	0.347	0.667	0.721	-0.055	0.029	0.062
	5	2	2.775	2.656	0.119	0.035	0.001	0.667	0.631	0.036	0.037	0.33
	5	3	2.775	2.76	0.014	0.031	0.641	0.667	0.702	-0.036	0.042	0.402
	5	4	2.775	2.826	-0.052	0.028	0.064	0.667	0.685	-0.018	0.03	0.542
Item4	2	1	2.634	2.647	-0.014	0.038	0.718	0.686	0.667	0.019	0.033	0.567
	3	1	2.437	2.647	-0.21	0.048	0	0.707	0.667	0.04	0.043	0.349
	3	2	2.437	2.634	-0.196	0.05	0	0.707	0.686	0.021	0.044	0.633
	4	1	2.765	2.647	0.118	0.037	0.001	0.73	0.667	0.064	0.029	0.027
	4	2	2.765	2.634	0.132	0.037	0	0.73	0.686	0.045	0.03	0.138
	4	3	2.765	2.437	0.328	0.044	0	0.73	0.707	0.024	0.039	0.54
	5	1	2.785	2.647	0.138	0.036	0	0.782	0.667	0.115	0.035	0.001
	5	2	2.785	2.634	0.152	0.037	0	0.782	0.686	0.096	0.036	0.008
	5	3	2.785	2.437	0.348	0.042	0	0.782	0.707	0.075	0.044	0.091
	5	4	2.785	2.765	0.02	0.026	0.44	0.782	0.73	0.051	0.028	0.068
Item5	2	1	2.539	2.669	-0.129	0.043	0.003	0.696	0.668	0.028	0.034	0.411
	3	1	2.404	2.669	-0.265	0.048	0	0.683	0.668	0.015	0.051	0.765
	3	2	2.404	2.539	-0.135	0.051	0.009	0.683	0.696	-0.013	0.051	0.804
	4	1	2.575	2.669	-0.093	0.042	0.027	0.624	0.668	-0.044	0.034	0.202
	4	2	2.575	2.539	0.036	0.04	0.363	0.624	0.696	-0.072	0.036	0.047
	4	3	2.575	2.404	0.172	0.049	0.001	0.624	0.683	-0.059	0.054	0.27
	5	1	2.388	2.669	-0.281	0.039	0	0.544	0.668	-0.124	0.041	0.003
	5	2	2.388	2.539	-0.152	0.041	0	0.544	0.696	-0.152	0.042	0
	5	3	2.388	2.404	-0.016	0.044	0.71	0.544	0.683	-0.14	0.058	0.016
	5	4	2.388	2.575	-0.188	0.039	0	0.544	0.624	-0.081	0.041	0.049

DEĞİŞEN MADDE FONKSİYONUNUNUN TESPİTİNDE GPCMLasso ve HİZALAMA YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Öz

Bu çalışmanın amacı son yıllarda değişen madde fonksiyonunun (DMF) belirlenmesi için önerilen GPCMLasso ve hizalama yöntemlerinin gerçek veri üzerinde karşılaştırılarak güçlü ve zayıf yönlerinin belirlenmesidir. Bu amaçla Uluslararası Öğrenci Değerlendirme Programı (PISA) 2018 uygulamasının Bilgi, İletişim ve Teknolojilere (BİT) İlgili anketinin maddelerinin değişen madde fonksiyonu, uygulamaya katılan Doğu Avrupa ve Orta Asya ülkeleri arasında incelenmiştir. PISA 2018 uygulaması BİT anketine Türkiye, Bulgaristan, Hırvatistan, Gürcistan ve Kazakistan katılmış olup analizler 15 yaş gurubunda 29,277 öğrencinin verileri ile yapılmıştır. Hizalama yaklaşımına göre 2, 3, 4 ve 6 numaralı maddelerin faktör yükleri; 2 ve 6 numaralı maddelerin ise faktör kesenleri tüm ülkeler arasında değişmez bulunmuştur. GPCMLasso sonuçlarına göre ölçme aracının tüm maddelerinin ülkeler arasında hem DMF ve hem de değişen adım fonksiyonu (DAF) içerdiği belirlenmiştir. Sonuç olarak DMF tespitinde yöntemler arasındaki uyum %64 olarak hesaplanmıştır. GPCMLasso yönteminin hizalama yöntemine nazaran DMF tespitinde daha katı bir yöntem olduğu ortaya konmuştur.

Anahtar kelimeler: madde işlev farklılığı, değişen adım fonksiyonu, düzenleme, makine öğrenmesi, PISA 2018

GİRİŞ

Geçerlik, bir özelliğin tam ve doğru olarak ölçülebilme derecesidir (Mellor, 1995). Bazı durumlarda bireylerin ölçmek istediğimiz özelliklerine dil, cinsiyet, etnik ve kültürel köken, sosyo-ekonomik düzey gibi demografik veya problem çözme becerisi, testi yanıtlama davranışı gibi bilişsel özellikler karışabilir. Bu arka plan değişkenlerinin test puanlarına karışması testi alan belirli gruplar için altta yatan gizil özelliğin hatalı bir şekilde belirlenmesine yol açar. Bu durum geçerlik üzerinde bir tehdittir ve test puanlarının yanlı olmasına neden olur (Penfield ve Camilli, 2006). Test puanlarının yanlılığı istatistiksel olarak değişen madde fonksiyonu (DMF) analizleri ile test edilir. DMF, aynı yetenek düzeyinde olan, fakat farklı arka plan özelliklerine sahip olan bireylerin test maddelerine doğru yanıt verme olasılıklarının farklılaşmasıdır (Penfield ve Camilli, 2006). Alanyazında DMF farklı ölçme kuramlarına dayalı birçok farklı teknik ile test edilmektedir. Bu teknikleri geleneksel (Chang ve diğerleri, 1996; Choi ve diğerleri, 2011; Holland ve Thayer, 1988; Swaminathan ve Rogers, 1990; Thissen ve diğerleri, 1993) ve yeni yaklaşımlar (Belzak, 2023; Cohen ve Bolt, 2005; De Ayala vd., 2002; Muthén ve Asparouhov, 2010; Schauburger ve Mair, 2020) başlıkları altında sınıflandırmak mümkündür.

Mantel Hanzel, SIBTEST gibi geleneksel yaklaşımların çoğu genellikle iki ya da çok kategorili olan tek bir değişkene göre DMF belirlemekte olup yaş, yanıtlama süresi gibi sürekli değişkenler için DMF analizleri yapılamamaktadır. Ayrıca tek bir değişkene göre analizler yürütüldüğünde katılımcıların birden fazla arka plan özelliğinin birlikte DMF kaynağı olabileceği durumu göz ardı edilmektedir (Belzak, 2023). Ayrıca bu tekniklerde gruplar arasında parametre karşılaştırmaları yapılırken tam sıfır kısıtlamaları (yani gruplar arasında parametrelerin tam eşitliği) kullanılır (Muthén ve Asparouhov, 2012). Özellikle bu durum, çoklu grup karşılaştırmalarında ($n \geq 10$) ilgili madde sınırlı sayıda grup arasında güçlü düzeyde DMF içerdiğinde, maddenin tüm gruplar için yanlı görünmesine neden olabilir (Van de Vijver ve diğerleri., 2019). Bir diğer sorun ise birden fazla olası DMF kaynağının test yanıtlarını ve puanları çeşitli doğrusal, doğrusal olmayan ve hatta parametrik olmayan yollarla etkileyebilmesi ve bunların birçoğunun bu yöntemlerle tespit edilememesidir (Belzak, 2023; Schauburger ve Mair, 2020). Geleneksel yöntemlerle DMF tespitinin sözü geçen sınırlılıklarını ortadan kaldırmak için alanyazında yaklaşık ölçme değişmezliği yöntemleri altında genellikle yapısal eşitlik modellemesi temelli yöntemler (Muthén ve Asparouhov, 2012; Muthén ve Asparouhov, 2014) ve Makine Öğrenmesi (MÖ) yaklaşımları olarak bilinen çeşitli düzenleme veya yanlılık düzeltme yöntemleri önerilmiştir (Belzak ve Bauer, 2020; Liang ve Jacobucci, 2020; Tutz ve Schauburger, 2015; Schauburger ve Tutz, 2016; Strobl vd., 2015). Yaklaşık ölçme değişmezliği yöntemleri özellikle çoklu grup karşılaştırmalarının yapıldığı geniş ölçekli değerlendirmelerde test puanlarının gruplar arasında karşılaştırılabilir olabileceği en uygun ölçme modelini belirlemeye odaklanırken (Muthén ve Asparouhov, 2012); MÖ yöntemleri, DMF'nin doğrusal, doğrusal olmayan ve parametrik olmayan formatlarda birden fazla değişkenin birlikte değerlendirilmesine odaklanmıştır (Belzak, 2023).

Sonuç olarak, DMF önemli bir geçerlilik sorunudur (Xu ve Tracey, 2017) ve test puanlarına dayalı yüksek riskli kararların alındığı uygulamalarda test puanlarının geçerliğinin ortaya konması gerekmektedir. Bunun için var olan birçok tekniğin karşılaştırmalı olarak incelenip güçlü ve zayıf yönlerinin ortaya konması önemlidir. Geleneksel

DMF belirleme yöntemlerinin aksine yeni yaklaşımlar hem geleneksel yöntemlerin gerektirdiği katı varsayımları esnetmesi hem de DMF'nin birden fazla kaynağı olması durumunda bunların arasındaki doğrusal olan veya olmayan ilişkileri keşfederek DMF'nin asıl kaynağını belirlemesi açısından değerlidir. Bu bağlamda bu çalışmanın problemi, DMF belirlemede yeni yaklaşımlardan olan yapısal eşitlik modellemesi temelli hizalama yaklaşımı ile makine öğrenmesi temelli lasso cezalandırma yaklaşımının PISA 2018 BİT'e ilgi anketi maddeleri üzerinde ülkeler arası karşılaştırma çalışmasının yapılması ve yöntemlerin güçlü ve zayıf yönlerinin belirlenmesidir.

YÖNTEM

Araştırma Modeli

Betimsel çalışmalar bireylerin, ortamın, koşulların veya olayların mevcut durumunu tanımlamak ve yorumlamak amacını taşır ve yalnızca ilgilenilen olguyu inceler; bireyleri, koşulları veya olayları manipüle etmeye yönelik hiçbir girişimde bulunulmaz (Mertler, 2014). Bu araştırmada PISA 2018 BİT anketi kullanılarak hizalama ve GPCMLasso DMF belirleme yöntemlerinin karşılaştırılması yapıldığından betimsel bir araştırma yaklaşımı benimsenmiştir

Örnekleme

OECD (2021) raporunda Doğu Avrupa ve Orta Asya (EECA) ülkelerinde eğitim ve öğrenme fırsatlarının eksikliğinin çocukların geleceğini tehdit etmeye ettiği ve eğitimde reform yapmak için EECA ülkelerinin kendi eğitim sistemlerinin performansını anlamaları ve sonuçlarını diğer ülkelerinkilerle karşılaştırmaları gerektiği vurgulanmaktadır. Bu bağlamda analizler Doğu Avrupa ve Orta Asya (EECA) ülkelerinin katılımcıları ile gerçekleştirilmiştir. PISA-EECA ülkeleri 2000 yılından bu yana PISA uygulamalarına katılımlarını sürekli artırmış olup 2018 uygulamasına Bakü (Azerbaycan), Belarus, Bulgaristan, Hırvatistan, Gürcistan, Kazakistan, Moldova, Romanya, Türkiye ve Ukrayna olmak üzere 10 ülke katılım göstermiştir. BİT'e ilgi anketine ise Bulgaristan, Hırvatistan, Gürcistan, Kazakistan ve Türkiye olmak üzere beş ülke katılmıştır (OECD, 2021). Tablo 1'de BİT'e ilgi anketini uygulayan beş ülkenin istatistiklerine yer verilmiştir.

Tablo 1. Çalışma Grubuna İlişkin Frekans ve Yüzde Değerleri

	Ülke	Cinsiyet	f	%	N
BRG	Bulgaristan	Kız	1276	49.6	2574
		Erkek	1298	50.4	
GEO	Gürcistan	Kız	1336	51.1	2613
		Erkek	1277	48.9	
HRV	Hırvatistan	Kız	2508	51.9	4835
		Erkek	2327	48.1	
KAZ	Kazakistan	Kız	6724	48.8	13775
		Erkek	7051	51.2	
TUR	Türkiye	Kız	2736	49.9	5480
		Erkek	2744	50.1	

Veri Toplama Aracı

Bu çalışmada PISA 2018 BİT'e ilgi anketinin verileri kullanılmıştır. BİT'e ilgi anketi 1 = "kesinlikle katılmıyorum", 2 = "katılmıyorum", 3 = "katılıyorum", 4 = "kesinlikle katılıyorum" olmak üzere dört tepki kategorili Likert türündedir. Tablo 2'de her bir ülke için BİT'e ilgi anketinin betimsel istatistikleri, Cronbach alfa güvenilirlikleri ve doğrulayıcı faktör analizi uyum değerleri sunulmuştur. Her bir ülke için BİT'e ilgi anketinin geçerli ve güvenilir olduğu görülmüştür.

Tablo 2. BİT'e ilgi Anketinin Betimsel İstatistikleri ve DFA Model Uyum Değerleri

Ülke	(ort±ss)	α	CFI	TLI	SRMR
BRG	4.33±1.1	.90	.96	.97	.050
GEO	4.48±1.1	.90	.98	.97	.022
HRV	3.34±0.8	.82	.97	.95	.038
KAZ	3.57±0.9	.85	.96	.94	.035
TUR	4.29±1.1	.88	.97	.98	.045

Veri Analizi

Hizalama Yöntemi

Hizalama yöntemi, gruplar arasında ölçme parametrelerinin minimal düzeyde farklılaşmasına izin veren bir model üreterek ölçme değişmezliğini test eder. Yöntem, ölçme değişmezliğini varsaymayıp en ideal ölçme değişmezliği örüntüsünü keşfederek, her bir grup için faktör ortalamaları ve varyanslarını kestirebilir. Bu yöntemin en büyük avantajı yapısal modele dayanması ve çok sayıda grup için modelleri uygun şekilde tahmin edebilmesidir. Bunu yaparken de model uyumundan ödün vermez. Hizalama yöntemi, değişmez olmayan (değişen) madde parametrelerinin sayısını ve bozulan ölçme değişmezliğinin miktarını minimumda tutulabilen modeli tahmin ederek faktör yüklerini (λ_g), kesenlerini (v_g), faktör ortalamalarını (α_g) ve varyanslarını (ψ_g) kestirebilir (Asparouhov ve Muthén, 2014).

Hizalama yöntemi ile ölçme değişmezliği iki adımda test edilir. İlk adımda faktör yükü ve kesenlerin gruplar arasında serbest olduğu ve tüm gruplarda faktör ortalamalarının 0'a, varyanslarının ise 1'e sabitlendiği yapısal model kestirilir. Temel model "M₀" olarak adlandırılan bu yapısal model, gruplar arası parametre kısıtları içermediğinden, çoklu grup faktör analiz modelleri içerisinde en iyi uyum gösteren modeldir. İkinci adımda hizalama optimizasyonu yapılır. Bu aşamada faktör ortalamaları ve varyansları serbest bırakılır ve bir basitlik fonksiyonu (Eşitlik 1'de sunulmuştur) aracılığı ile ölçme değişmezliği miktarını bozan grup sayısını en aza indirgeyen faktör ortalamaları ve varyansları hesaplanır (Asparouhov ve Muthén, 2014.). Basitlik fonksiyonu aracılığı ile tanımlanmamış model, faktör ortalamalarının ve varyanslarının yapısal modele eklenmesiyle tanımlanmış olur.

$$F = \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{pg_1} - \lambda_{pg_2}) + \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(v_{pg_1} - v_{pg_2}) \quad (1)$$

$$w_{g_1, g_2} = \sqrt{N_{g_1} N_{g_2}} \quad (2)$$

w:factor weight,

N: sample size of the group

Önerilen son hizalanmış model (M_1), M_0 modeliyle aynı uyuma sahiptir. Yani hizalanmış model değişmezlik miktarını en aza indirmeye çalışmasına rağmen, model uyumundan ödün vermez. Hizalama optimizasyonunun model uyumunu değerlendirmek için Asparouhov ve Muthén, (2014) istatistiksel olarak değişmezliği bozan madde sayısının, toplamın %25'ini aşmaması gerektiğini; Flake ve McCoach (2018) ise %29'u aşmaması gerektiğini belirtmişlerdir. Hizalama modelinin uyumunu değerlendirmek için kullanılacak bir diğer ölçüt R^2 değişmezlik etki büyüklüğü ölçüsüdür. R^2 madde parametre kestirimindeki değişkenliğin ne kadarının grupların faktör ortalamaları ve varyansları ile açıklanabildiğini gösterir. R^2 değerinin 1 civarında olması tam değişmezliği gösterir, çünkü bu durumda madde parametrelerindeki değişkenlik tamamen grup ortalama farklılıkları ile açıklanır. 0 civarında bir R^2 ise grup ortalama farklılıklarının madde parametrelerindeki değişkenliğin hiçbirini açıklamadığını gösterir (Muthén ve Asparouhov, 2014). Bu çalışmada hizalama yöntemine dayalı DMF analizleri Mplus 8.0 (Muthén, B., & Muthén, 2017) ile yapılmıştır.

GPCMlasso

Lasso düzenleştirmesinin uygulandığı genelleştirilmiş kısmi kredi modeli (GPCMlasso), genelleştirilmiş kısmi kredi modeli ile DMF tespiti yapmak için parametre seçiminde cezalandırılmış olabilirlik tahmini kullanan birleşik bir model oluşturur (Schauberger & Mair, 2020). Bu yöntemin büyük avantajı, birkaç değişkenin aynı anda ele alınabilmesi ve DMF tespiti için hem sürekli hem de kategorik değişkenlerin kullanılabilmesidir. GPCMlasso ile tek biçimli DMF ülke (G) değişkenine göre eşitlik 3'de gösterildiği şekilde hesaplanır.

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \beta_i = [\theta_p + x_p^T \alpha - \delta_{ir} - (\gamma_i \times G)] \quad (3)$$

Bu modelde γ_i ülke değişkeni için i maddesine ilişkin DMF parametrelerini temsil eder. Lasso düzenleştirmesi uygulandıktan sonra bu parametreler sıfıra eşit değilse, ilgili madde tek biçimli DMF içerir. Parametre kestirimi için GPCMlasso modeli eşitlik 4'de gösterilen lasso düzenleştirmesinin log-olabilirlik fonksiyonunu çözer.

$$\ell_p(\theta, \alpha, \delta, \beta, \gamma) = \ell(\theta, \alpha, \delta, \beta, \gamma) - \lambda \sum_{i=1}^l \sum_{j=1}^m w_{ij} |\gamma_{ij}| \quad (4)$$

Burada $\ell(\theta, \alpha, \delta, \beta, \gamma)$ log-olabilirlik fonksiyonunun temel versiyonunu, $\lambda \sum_{i=1}^l \sum_{j=1}^m w_{ij} |\gamma_{ij}|$ lasso düzenleme terimini, $\lambda \geq 0$ regresyon katsayıları (γ_{ij}) vektörüne uygulanan düzenleme derecesini kontrol eden ayar parametresini temsil eder. x_p , p bireyinin m boyutlu kovaryant vektörünü (G), θ ise gizil değişkeni temsil eder. γ_i , DMF parametrelerinin yanısıra δ_{ij} , i maddesinin madde adım parametrelerini, α , temel etkisini ve β_i ayırt edicilik parametresini temsil eder. Penfield (2007) ve Penfield ve diğerleri (2009), çok kategorili maddelerde DMF analizinin aynı değişen adım fonksiyonu (DAF) olgusundan da zarar görebileceğini belirtmiştir. DAF, farklı ortak değişken değerlerine göre çok kategorili bir maddenin her adımının düzeyindeki farklılıkları ifade eder. Bu nedenle bu çalışmada DMF ile birlikte DAF da incelenmiştir. Optimal ayar parametresinin (λ) belirlenmesinde BIC model seçim kriteri kullanılmıştır. GPCMLasso modeli için BIC, eşitlik 5'deki gibi hesaplanır.

$$BIC(\lambda) = -2L\lambda(\cdot) + df(\lambda) \log(n) \quad (5)$$

Bu çalışmada GPCMLasso yöntemine dayalı DMF analizleri R yazılımındaki GPCMLasso paketi (R Core Team, 2018) ile yapılmıştır.

BULGULAR

Bulgular, önce hizalama yöntemi bulguları ardından GPCMLasso bulguları olmak üzere iki aşamada sunulmaktadır.

Hizalama Yöntemi Bulguları

Hizalama analizleri yapılmadan önce çoklu grup doğrulayıcı faktör analizi (ÇGDFA) uygulanmıştır. Tablo 3' de, ÇGDFA uygulaması sonucunda elde edilen yapısal, metrik ve skalar değişmezlik seviyesine ilişkin uyum değerlerine yer verilmiştir. Ki-kare fark testleri, yapısal değişmezlikten metrik değişmezliğe ve metrik değişmezlikten skalar değişmezliğe geçişte, model uyumunun önemli ölçüde bozulduğunu göstermektedir. Alternatif uyum iyiliği indekslerindeki değişim incelendiğinde, hem yapısaldan metriğe hem de metriktan skalara geçişte CFI, RMSEA ve SRMR farkının kabul değerleri aştığı görülmektedir (Sokolov, 2019). Bu bulgu, altı maddeden oluşan BİT'e ilgi anketinin faktör ortalamalarının beş ülke arasında karşılaştırılabilir olmadığını göstermektedir.

Tablo 3. ÇGDFA model veri uyumu

Model	χ^2	df	p	RMSEA	CFI	SRMR
Yapısal	269.557	45	<.001	.095 [.085-.106]	.972	.031
Zayıf	390.661	65	<.001	.096 [.087-.105]	.960	.080
Güçlü	838.929	85	<.001	.127 [.119-.135]	.907	.078

Tablo 4'de beş ülke arasında 6 maddelik BİT'e ilgi anketinin her bir maddesinin kesen ve yük değerlerinin değişmezliğine ilişkin bulgular sunulmuştur. Parantez, faktör yükleri ve kesenlerinin ilgili ülke için anlamlı olarak

farklılaştığı anlamına gelmektedir. *BİT'e ilgi* maddelerinin faktör yükleri madde 2,3,4 ve 6 için ülkeler arasında değişmezdir. Öte yandan madde 1 için Gürcistan ve Kazakistan'ın faktör yükü anlamlı olarak farklılaşmıştır. Ek 1'de faktör yükleri ve kesenlerine göre değişmezliğin sağlanmadığı maddeler için, ülkeler arasında ikili karşılaştırma sonuçlarına yer verilmiştir. Buna göre Gürcistan'ın faktör yükü Türkiye ve Hırvatistan'dan; Kazakistan'ın faktör yükü ise tüm ülkelerden anlamlı derecede farklıdır. Madde 5 için ise Kazakistan'ın faktör yükü anlamlı olarak farklılaşmıştır. İkili karşılaştırma sonuçlarına göre bu farkın Türkiye ve Bulgaristan arasında olduğu görülmektedir. Genel olarak, *BİT'e ilgi* maddelerinin ülkeler arasında faktör yüklerinin %3'ü (≤ 25) (Muthén ve Asparouhov, 2014) gruplar arasında değişmezliği bozmaktadır. Buna göre ülkeler arasında faktör yükleri karşılaştırılabilir. Dolayısıyla *BİT'e ilgi* ölçeğinin metrik değişmezliği sağladığı söylenebilir. *BİT'e ilgi* maddelerinin ülkeler arasında faktör kesenleri incelendiğinde ise madde 2 ve madde 4'ün kesenlerinin ülkeler arasında değişmez olduğu görülmektedir. Öte yandan madde 1 için Bulgaristan, Hırvatistan ve Gürcistan; madde 3 için Bulgaristan, Hırvatistan, Gürcistan ve Kazakistan; madde 4 için Bulgaristan ve Türkiye; Madde 5 için ise Bulgaristan, Türkiye ve Gürcistan arasında ölçme değişmezliği sağlanmıştır. *BİT'e ilgi* maddelerinin ülkeler arasında faktör kesenleri incelendiğinde ise faktör kesenlerinin %27'sinin (≥ 25 , Muthén ve Asparouhov, 2014) gruplar arasında değişmezliği bozduğu görülmektedir. Buna göre *BİT'e ilgi* ölçeği skalar değişmezliği sağlamamaktadır. Dolayısıyla ülkeler arasında ölçek toplam puanına göre faktör ortalamaları karşılaştırılmaz.

Tablo 4. Faktör yükleri ve kesenlerine ilişkin hizalama bulguları

Madde	Kesen	Değişmezliği Bozan Parametre Sayısı	Faktör Yükü	Değişmezliği Bozan Parametre Sayısı
13Q01 (madde1)	1 (2) 3 4 (5)	2	1 2 3 (4) (5)	2
13Q04 (madde2)	1 2 3 4 5	0	1 2 3 4 5	0
13Q05 (madde3)	1 (2) 3 4 5	1	1 2 3 4 5	0
13Q11 (madde4)	1 2 (3) (4) (5)	3	1 2 3 4 5	0
13Q12 (madde5)	1 2 (3) 4 (5)	2	1 2 3 4 (5)	1
13Q13 (madde6)	1 2 3 4 5	0	1 2 3 4 5	0
Toplam		8(%26.6)		3 (%10)

*BGR=1, TUR=2, HRV=3, GEO=4, KAZ=5

Tablo 5'de hizalama yöntemiyle tahmin edilen faktör ortalamaları sunulmuştur. Buna göre, Hırvatistan'ın en yüksek, Kazakistan'ın en düşük faktör ortalamasına sahip olduğu görülmüştür. Hırvatistan ile diğer ülkelerin arasında faktör ortalamaları birbirlerinden anlamlı ölçüde farklılaşmaktadır. Benzer şekilde Kazakistan ve Türkiye'nin faktör ortalamaları Bulgaristan ve Gürcistan'dan anlamlı derecede yüksektir.

Tablo 5. Ülkeler Arası Faktör Ortalamalarının Karşılaştırılması

Sıra	Grup	Faktör Ortalaması	İstatistiksel olarak Düşük Faktör Ortalamalarına Sahip Gruplar
1	3	0.431	5 2 1 4
2	5	0.169	1 4
3	2	0.130	1 4
4	1	0.000	
5	4	-0.078	

*BGR=1, TUR=2, HRV=3, GEO=4, KAZ=5

GPCMlasso Bulguları

Tablo 6'da BIC kriterlerine göre ülkeler arasında DMF'yi tespit etmeye yönelik GPCMlasso modelinin sonuçları sunulmuştur.

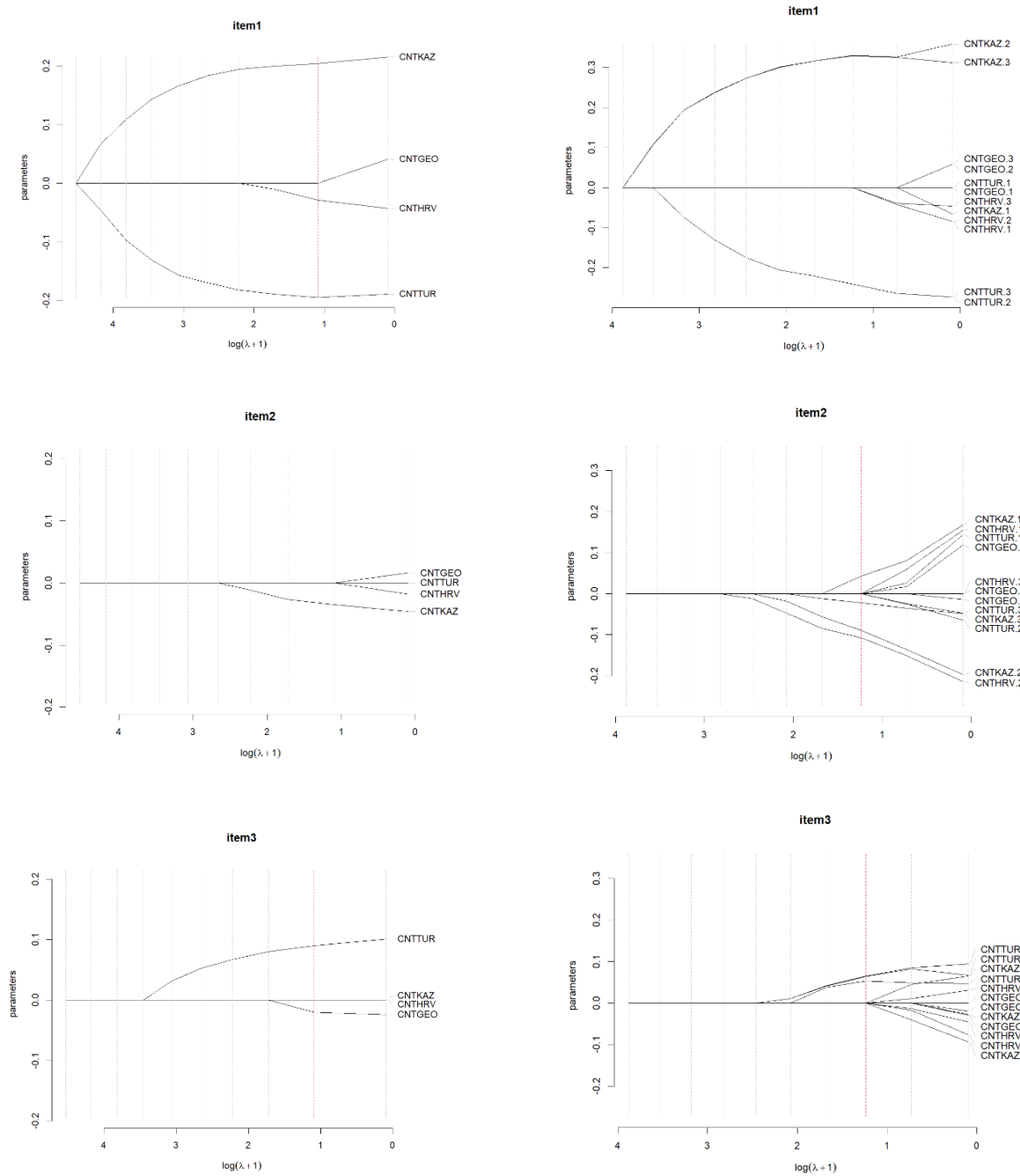
Tablo 6. GPCMlasso Katsayılarına İlişkin Değişkenlere Göre DMF Sonuçları

Madde Kodu	GEO	HRV	KAZ	TUR
13Q01 (madde1)	0.000	-0.029	0.204	-0.195
13Q04 (madde2)	0.000	0	-0.035	0
13Q05 (madde3)	-0.02	0	0	0.09
13Q11 (madde4)	-0.084	0.111	-0.117	0
13Q12 (madde5)	0	0.114	0.147	0.042
13Q13 (madde6)	0	-0.016	0	0

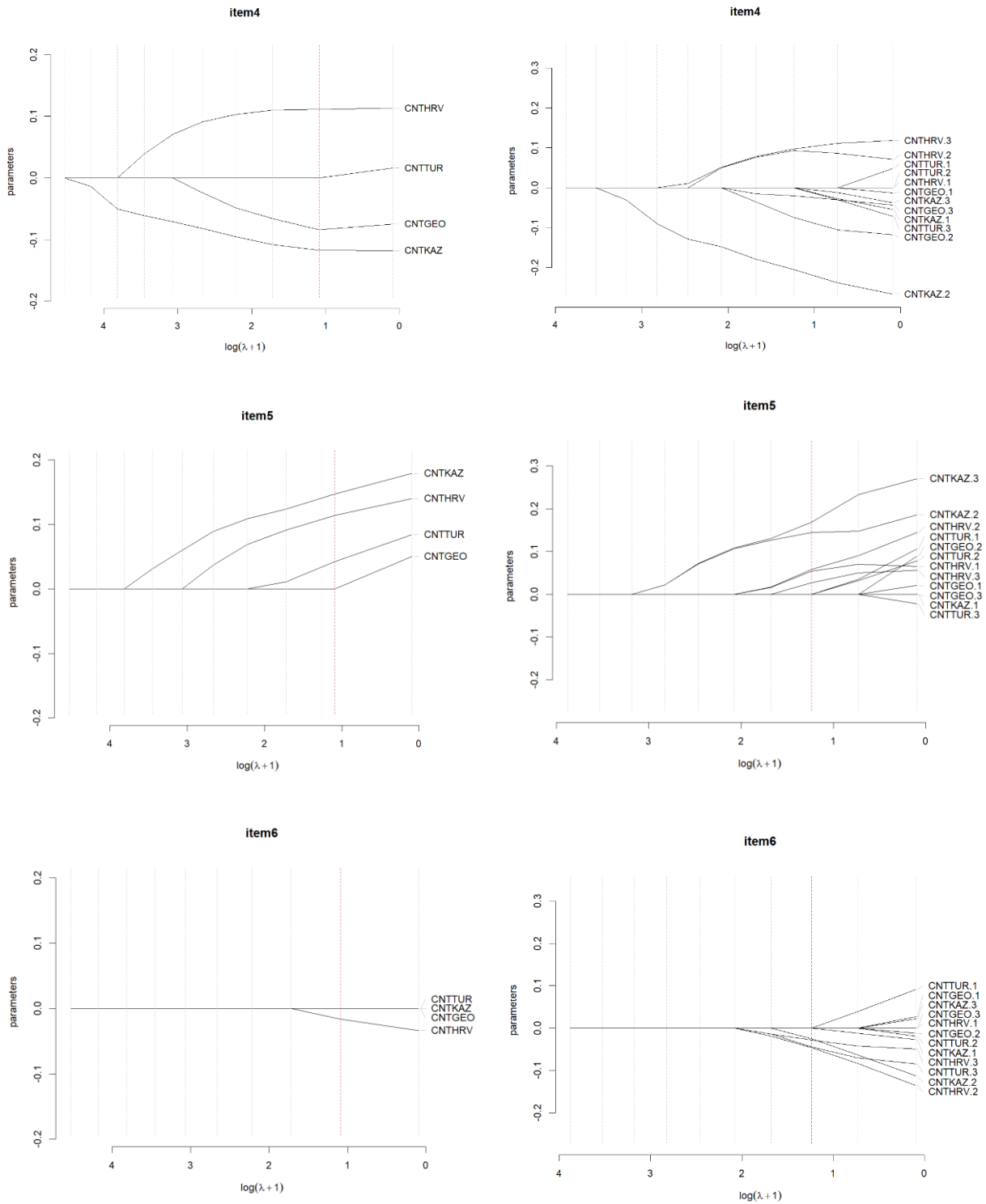
Bulgaristan referans ülke olarak alındığından ülke değişkenine ilişkin tüm sonuçlar, her ülkenin Bulgaristan ile ikili karşılaştırmasının sonuçlarını yansıtmaktadır. Lasso katsayısının sıfır olmadığı durumlarda ilgili madde grup değişkenine göre tek biçimli DMF içerir. Lasso katsayısının işareti, grup değişkeni ile ilgili maddenin arasındaki ilişkinin yönünü temsil eder. Pozitif katsayı, maddenin odak grubu lehine, negatif katsayı ise maddenin referans grup lehine yanlı olduğunu gösterir. Katsayının büyüklüğü aynı zamanda DMF etkisinin gücü veya büyüklüğü hakkında bilgi sağlar. Katsayının mutlak değeri ne kadar büyük olursa DMF etkisi de o kadar güçlü olur. Şekil 1'de ceza terimi (penalty term) uygulandığında λ ayarlama parametresi boyunca her bir madde için DMF parametrelerinin izlediği parametre yollarına ilişkin grafikler sunulmuştur. Yollar her madde için ayrı ayrı çizilir. Kırmızı kesikli çizgiler BIC'e göre en uygun modeli temsil eder. Bu şekilde, belirli bir değişkenin parametre yolu dikey kırmızı kesikli çizgiyi geçiyorsa, ilgili madde DMF içerir. Şekil 1'in sol tarafında ilgili maddenin ülke değişkenine göre değişen madde fonksiyonu, sağ tarafında ise DAF bulguları sunulmuştur. GPCMlasso bulguları sunulurken Tablo 6 ve Şekil 1 birlikte değerlendirilmiştir.

BiT'e ilgi alt ölçeğinin tüm maddeleri ülke değişkenine göre DMF içermektedir. Madde 1, Bulgaristan ile Kazakistan arasında Kazakistan lehine; Bulgaristan ile Hırvatistan arasında Bulgaristan lehine ve Bulgaristan ile Türkiye arasında Bulgaristan lehine DMF vardır. Değişen adım fonksiyonları incelendiğinde madde 1'in ikinci ve üçüncü adım parametresi Bulgaristan ile Kazakistan arasında Kazakistan'ın lehine; Bulgaristan ile Türkiye arasında Bulgaristan lehine farklı fonksiyonlaşmıştır. Öte yandan madde 1 Bulgaristan ile Hırvatistan arasında DMF göstermiş olsa da DAF göstermemektedir. Bunun nedeni DMF'nin ihmal edilebilir düzeyde olmasından kaynaklanabilir. Madde 2 sadece Bulgaristan ile Kazakistan arasında Bulgaristan lehine DMF göstermiştir. Öte yandan DAF sonuçlarına göre madde 2'nin birinci adım parametresinde Kazakistan, ikinci adım parametresinde ise Bulgaristan lehine DAF görülmüştür. Ayrıca Bulgaristan ile Kazakistan arasında DMF olmamasına rağmen maddenin ikinci adım parametresi bu iki ülke arasında Bulgaristan lehine DAF göstermiştir. Madde 3 Bulgaristan ile Türkiye arasında Türkiye lehine hem DMF hem de bütün madde adım parametreleri için DAF içermektedir. Öte yandan madde 3 Bulgaristan ile Gürcistan arasında Bulgaristan lehine ihmal edilebilir düzeyde DMF içermekte olup DAF göstermemiştir. Madde 4, Bulgaristan ile Hırvatistan arasında Hırvatistan lehine DMF göstermekte olup

ikinci ve üçüncü adım parametresi için de DAF göstermiştir. Madde 4 aynı zamanda Gürcistan ve kazakistan için de Bulgaristan lehine DMF içermektedir. Her iki ülke için de ikinci adım fonksiyonunda DAF vardır. Madde 5, Bulgaristan ile Kazakistan arasında Kazakistan lehine DMF; ikinci ve üçüncü adım parametresi için DAF içermektedir. Bulgaristan ile Hırvatistan arasında Hırvatistan lehinde DMF ve tüm adım parametreleri için DAF içermektedir. Bulgaristan ile Türkiye arasında Türkiye lehine DMF ve birinci ve ikinci adım parametrelerinde DAF içermektedir. Madde 6 ise Bulgaristan ile Hırvatistan arasında Bulgaristan lehine DMF, ikinci ve üçüncü adım parametrelerinde DAF içermektedir.



Şekil 1. GPCMLasso'ya Ait Tüm DMF ve DAF Parametrelerinin Katsayı Yolları.



Şekil 1 (Devam). GPCmlasso'ya Ait Tüm DMF ve DAF Parametrelerinin Katsayı Yolları.

Tablo 7’de Hizalama ve GPCMLasso bulgularına göre referans ülke Bulgaristan ile diğer ülkeler arasındaki DMF sonuçları sunulmuştur. DMF tespitinde yöntemler arasındaki uyum %64 olarak hesaplanmıştır. Hizalama yöntemi ile madde 2 ve 6 hiçbir ülke arasında DMF içermezken; GPCM lasso yöntemi ile madde 2 Bulgaristan ile Kazakistan, madde 6 Bulgaristan ile Hırvatistan arasında farklı fonksiyonlaşmıştır. Ancak bu maddelerde DMF düzeyi ihmal edilebilir olduğundan GPCMLasso yönteminin hizalama yöntemine nazaran DMF tespitinde daha hassas bir yöntem olduğu söylenebilir.

Tablo 7. GPCMLasso ve Hizalama Yöntemlerine Göre DMF Sonuçları

	Madde 1		Madde 2		Madde 3		Madde 4		Madde 5		Madde 6	
	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI	Alg.	GPCMI
BGR-TUR	+	+			+	+			+	+		
BGR-HRV		+					+	+	+	+		(+)
BGR-GR						+	+	+	+			
BGR-KAZ	+	+		(+)			+	+	+	+		

TARTIŞMA ve SONUÇ

Bu çalışmada, PISA 2018 BİT’e ilgi anket maddelerinin Doğu Avrupa ve Orta Asya ülkeleri arasında bir makine öğrenme algoritması olan GPCMLasso ve yapısal eşitlik modellemesi temelli hizalama yöntemine göre değişen madde fonksiyonu incelemesi yapılmıştır. Hizalama yaklaşımına göre 2, 3, 4 ve 6 numaralı maddelerin faktör yükleri; 2 ve 6 numaralı maddelerin ise faktör kesenleri tüm ülkeler arasında değişmezdir. Değişmezliği bozan parametre oranı faktör yükleri arasında %3 iken faktör kesenleri arasında %27 (≥ 25 , Muthén ve Asparouhov, 2014) olduğundan BİT’e ilgi anketinin skalar değişmezliği sağlamadığı sonucuna varılmıştır. Bu sonuca göre BİT’e ilgi ölçeğinin ülkeler arasında faktör ortalamaları karşılaştırılmaz. Alanyazında PISA-BİT anketinin ülkeler arasında ölçme değişmezliğinin incelendiği birkaç çalışma mevcuttur. Odell ve diğerleri (2021), 47 ülke üzerinde PISA 2015 BİT anketinin ölçme değişmezliğini incelemek için hizalama tekniğini kullanmış ve ülkeler arasında BİT anketinden elde edilen puan ortalamalarının karşılaştırılmasının mümkün olmadığını belirtmişlerdir. Bu bulgu bu çalışmadan elde edilen araştırma bulguları ile tutarlıdır. Öte yandan Ma ve Qin (2021) PISA 2018 BİT anketinin 16 ülke arasında ölçme değişmezliğini çoklu grup doğrulayıcı faktör analizi ile test etmiş ve artık (katı) değişmezliğin sağlandığını ortaya koymuştur.

GPCMLasso sonuçları, PISA 2018 BİT’e ilgi anketinin tüm maddelerinin ülkeler arasında DMF ve DAF içerdiğini göstermiştir. Sırasıyla referans ülke Bulgaristanla Gürcistan arasında iki madde (madde 3, 4); Hırvatistan arasında dört madde (madde 1, 4, 5, 6); Kazakistan arasında dört madde (madde 1, 2, 4, 6) ve Türkiye arasında üç madde (madde 1, 3, 5) DMF içermektedir. Ayrıca GPCMLasso, değişen adım fonksiyonu analizi yaparak çok kategorili maddelerin kategori parametrelerinin gruplar arasında farklı fonksiyonlaşma durumuna ilişkin bilgi sağlar. Bu çalışmada genellikle referans ülke Bulgaristan ile DAF gösteren ülkelerin özellikle ikinci ve üçüncü kategori parametrelerinde farklı fonksiyonlaşmanın olduğu gözlenmiştir.

Sonuç olarak, her iki yöntem de PISA 2018 CIT'e ilgi anketinde ülke değişkenine göre DMF içeren maddeleri benzer şekilde tespit etmiştir. Öte yandan GPCMLasso, ihmal edilebilir miktarda DMF yakaladığı için DMF'yi tespit etmede daha katıdır. Hizalama yaklaşımı tüm ülkeler arasındaki ikili karşılaştırma sonuçlarını tek bir analizde sunarken, GPCMLasso yöntemi referans ülkeye göre ikili karşılaştırma sonuçlarını sunmaktadır. Ayrıca GPCMLasso yöntemi çok kategorili maddeler arasında hangi madde yanıt kategorisi düzeyinde grup farklılaşmasının olduğunu hesaplaması açısından öne çıkmaktadır.

ÖNERİLER

Bu araştırmada PISA 2018 BİT'e ilgi anketinin hizalama ve GPCMLasso yöntemleri ile DMF analizleri yürütülmüştür. Her iki yöntem de çok boyutlu yapılarda eş zamanlı analiz yapabilmektedir. Dolayısıyla sonraki çalışmalarda çok boyutlu yapılarda yöntemlerin performansı incelenebilir. Ayrıca bu çalışmada DMF tespitinde yeni yaklaşımlar olan yapısal eşitlik modellemesi temelli hizalama yaklaşımı ile makine öğrenmesi temelli GPCMLasso yöntemi karşılaştırılmıştır. Yöntemler arası karşılaştırmalar MTK temelli DMF belirleme yaklaşımları da dahil edilerek yapılabilir. Bu nedenle, hangi yöntemin belirli koşullar altında DMF'yi en doğru şekilde tanımlayabildiğini anlamak için farklı yaklaşımlara dayanan yöntemleri inceleyen bir simülasyon çalışması faydalı olabilir.

Etik Metni

Bu makalede dergi yazım kurallarına, yayın ilkelerine, araştırma ve yayın etiği kurallarına, dergi etik kurallarına uyulmuştur. Makale ile ilgili doğabilecek her türlü ihlallerde sorumluluk yazarlara aittir. Bu çalışma etik kurul gerektirmeyen bir çalışmadır.

Çıkar çatışması: Yazar herhangi bir çıkar çatışması beyan etmemiştir.

Finansal Destek: Bu çalışmada herhangi bir kurum veya kuruluştan finansal destek alınmamıştır.

Yazar Katkı Oranı: Bu çalışmada tek yazarın katkı oranı %100'dür.

KAYNAKÇA

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508. <https://doi.org/10.1080/10705511.2014.919210>
- Belzak, W. C. M., & Bauer, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychological Methods*, 25(6), 673–690. <https://doi.org/10.1037/met0000253>
- Belzak, W. C. (2023). The Multidimensionality of Measurement Bias in High-Stakes Testing: Using Machine Learning to Evaluate Complex Sources of Differential Item Functioning. *Educational Measurement: Issues and Practice*, 42(1), 24-33. <https://doi.org/10.1111/emip.12486>

- Chang, H. H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items an adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333–353. <https://doi.org/10.1111/j.1745-3984.1996.tb00496.x>
- Choi, S., Gibbons, L., & Crane, P. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(1), 1–30. <https://doi.org/10.18637/jss.v039.i08>
- Flake, J. K., & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1), 56–70. <https://doi.org/10.1080/10705511.2017.1374187>
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Lawrence Erlbaum Associates, Inc.
- Liang, X., & Jacobucci, R. (2020). Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 722–734. <https://doi.org/10.1080/10705511.2019.1693273>
- Ma, Y., & Qin, X. (2021). Measurement invariance of information, communication and technology (ICT) engagement and its relationship with student academic literacy: Evidence from PISA 2018. *Studies in Educational Evaluation*, 68, 100982. <https://doi.org/10.1016/j.stueduc.2021.100982>
- Mellor, L. T. (1995). *A comparison of four differential item functioning (DIF) methods for polytomously scored items* (Order No. 9534889). Available from ProQuest One Academic. (304235582). <http://proxy.libraries.smu.edu/login?url=https://www.proquest.com/dissertations-theses/comparison-four-differential-item-functioning-dif/docview/304235582/se->
- Mertler, C. A. (2014). *The Data-Driven Classroom: How do I use student data to improve my instruction?* (ASCD Arias). ASCD.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <https://doi.org/10.1037/a0026802>
- Odell, B., Gierl, M., & Cutumisu, M. (2021). Testing measurement invariance of PISA 2015 mathematics, science, and ICT scales using the alignment method. *Studies in Educational Evaluation*, 68, 100965. <https://doi.org/10.1016/j.stueduc.2020.100965>
- OECD, U. (2021). *Education in Eastern Europe and Central Asia: Findings from PISA*.
- Penfield, R. D., & Camilli, G. (2006). Differential Item Functioning and Item Bias. *Handbook of statistics*, 26, 125–167. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X)
- Penfield, R. D. (2007). Assessing differential step functioning in polytomous items using a common odds ratio estimator. *Journal of Educational Measurement*, 44(3), 187–210. <https://doi.org/10.1111/j.1745-3984.2007.00034.x>
- Penfield, R. D., Gattamorta, K., & Childs, R. A. (2009). An NCME instructional module on using differential step functioning to refine the analysis of DIF in polytomous items. *Educational Measurement: Issues and Practice*, 28(1), 38–49. <https://doi.org/10.1111/j.1745-3992.2009.01135.x>
-

- Schauberger, G., Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behav Res* 52, 279–294 (2020). <https://doi.org/10.3758/s13428-019-01224-2>
- Schauberger, G., & Tutz, G. (2016). Detection of differential item functioning in Rasch models by boosting techniques. *British Journal of Mathematical and Statistical Psychology*, 69(1), 80–103. <https://doi.org/10.1111/bmsp.12060>
- Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289–316. <https://doi.org/10.1007/s11336-013-9388-3>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In Holland, P. W., & Wainer, H. (Eds.) *Differential item functioning*, (pp. 67-113).
- Tutz, G., & Schauberger, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika*, 80(1), 21–43. <https://doi.org/10.1007/s11336-013-9377-6>
- Xu, H., & Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *The European Journal of Counselling Psychology*, 6(1), 75–82. <https://doi.org/10.5964/ejcop.v6i1.120>
- Van de Vijver, F. J., Avvisati, F., Davidov, E., Eid, M., Fox, J. P., Le Donne, N., ... & van de Schoot, R. (2019). *Invariance analyses in large-scale studies*. <https://doi.org/10.1787/19939019>

EKLER

Ek 1. Ülkeler arası ikili karşılaştırmalar

	Kesen							Faktör Yüğü				
	Grup	Grup	Deęer	Deęer	Fark	Std. H	p	Deęer	Deęer	Fark	Std. H	p
Madde1	2	1	2.677	2.418	0.258	0.046	0	0.742	0.624	0.118	0.048	0.015
	3	1	2.424	2.418	0.006	0.039	0.879	0.707	0.624	0.083	0.061	0.176
	3	2	2.424	2.677	-0.253	0.048	0	0.707	0.742	-0.035	0.058	0.551
	4	1	2.349	2.418	-0.069	0.046	0.136	0.545	0.624	-0.079	0.045	0.08
	4	2	2.349	2.677	-0.327	0.049	0	0.545	0.742	-0.197	0.047	0
	4	3	2.349	2.424	-0.075	0.047	0.115	0.545	0.707	-0.162	0.061	0.008
	5	1	2.135	2.418	-0.283	0.043	0	0.338	0.624	-0.286	0.051	0
	5	2	2.135	2.677	-0.542	0.048	0	0.338	0.742	-0.404	0.054	0
	5	3	2.135	2.424	-0.289	0.044	0	0.338	0.707	-0.369	0.066	0
	5	4	2.135	2.349	-0.215	0.045	0	0.338	0.545	-0.207	0.051	0
Madde3	2	1	2.656	2.8	-0.144	0.036	0	0.631	0.721	-0.09	0.034	0.008
	3	1	2.76	2.8	-0.039	0.034	0.248	0.702	0.721	-0.019	0.04	0.632
	3	2	2.76	2.656	0.105	0.041	0.011	0.702	0.631	0.071	0.048	0.14
	4	1	2.826	2.8	0.027	0.029	0.352	0.685	0.721	-0.036	0.026	0.169
	4	2	2.826	2.656	0.171	0.037	0	0.685	0.631	0.054	0.035	0.127
	4	3	2.826	2.76	0.066	0.035	0.056	0.685	0.702	-0.017	0.04	0.671
	5	1	2.775	2.8	-0.025	0.027	0.347	0.667	0.721	-0.055	0.029	0.062
	5	2	2.775	2.656	0.119	0.035	0.001	0.667	0.631	0.036	0.037	0.33
	5	3	2.775	2.76	0.014	0.031	0.641	0.667	0.702	-0.036	0.042	0.402
	5	4	2.775	2.826	-0.052	0.028	0.064	0.667	0.685	-0.018	0.03	0.542
Madde4	2	1	2.634	2.647	-0.014	0.038	0.718	0.686	0.667	0.019	0.033	0.567
	3	1	2.437	2.647	-0.21	0.048	0	0.707	0.667	0.04	0.043	0.349
	3	2	2.437	2.634	-0.196	0.05	0	0.707	0.686	0.021	0.044	0.633
	4	1	2.765	2.647	0.118	0.037	0.001	0.73	0.667	0.064	0.029	0.027
	4	2	2.765	2.634	0.132	0.037	0	0.73	0.686	0.045	0.03	0.138
	4	3	2.765	2.437	0.328	0.044	0	0.73	0.707	0.024	0.039	0.54
	5	1	2.785	2.647	0.138	0.036	0	0.782	0.667	0.115	0.035	0.001
	5	2	2.785	2.634	0.152	0.037	0	0.782	0.686	0.096	0.036	0.008
	5	3	2.785	2.437	0.348	0.042	0	0.782	0.707	0.075	0.044	0.091
	5	4	2.785	2.765	0.02	0.026	0.44	0.782	0.73	0.051	0.028	0.068
Madde5	2	1	2.539	2.669	-0.129	0.043	0.003	0.696	0.668	0.028	0.034	0.411
	3	1	2.404	2.669	-0.265	0.048	0	0.683	0.668	0.015	0.051	0.765
	3	2	2.404	2.539	-0.135	0.051	0.009	0.683	0.696	-0.013	0.051	0.804
	4	1	2.575	2.669	-0.093	0.042	0.027	0.624	0.668	-0.044	0.034	0.202
	4	2	2.575	2.539	0.036	0.04	0.363	0.624	0.696	-0.072	0.036	0.047
	4	3	2.575	2.404	0.172	0.049	0.001	0.624	0.683	-0.059	0.054	0.27
	5	1	2.388	2.669	-0.281	0.039	0	0.544	0.668	-0.124	0.041	0.003
	5	2	2.388	2.539	-0.152	0.041	0	0.544	0.696	-0.152	0.042	0
	5	3	2.388	2.404	-0.016	0.044	0.71	0.544	0.683	-0.14	0.058	0.016
	5	4	2.388	2.575	-0.188	0.039	0	0.544	0.624	-0.081	0.041	0.049