International Journal of Assessment Tools in Education
2022, Vol. 9, No. 3, 728–740
https://doi.org/10.21449/ijate.1098757
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

To what extent are item discrimination values realistic? A new index for two-dimensional structures

Abdullah Faruk Kilic [1,*], **Ibrahim Uysal** [2]

[1]Adıyaman University, Faculty of Education, Department of Educational Sciences, Division of Educational Measurement and Evaluation, Adıyaman, Türkiye
[2]Bolu Abant İzzet Baysal University, Faculty of Education, Department of Educational Sciences, Division of Educational Measurement and Evaluation, Bolu, Türkiye

ARTICLE HISTORY

Received: Apr. 05, 2022
Revised: July 18, 2022
Accepted: Sep. 01, 2022

Keywords:
Item discrimination,
Multi-dimensional structures,
Classical test theory,
Corrected item-total correlation.

Abstract: Most researchers investigate the corrected item-total correlation of items when analyzing item discrimination in multi-dimensional structures under the Classical Test Theory, which might lead to underestimating item discrimination, thereby removing items from the test. Researchers might investigate the corrected item-total correlation with the factors to which that item belongs; however, getting a general overview of the entire test is impossible. Based on this problem, this study aims to recommend a new index to investigate item discrimination in two-dimensional structures through a Monte Carlo simulation. The new item discrimination index is evaluated by identifying sample size, item discrimination value, inter-factor correlation, and the number of categories. Based upon the results of the study it can be claimed that the proposed item discrimination index proves acceptable performance for two-dimensional structures. Accordingly, using this new item discrimination index could be recommended to researchers when investigating item discrimination in two-dimensional structures.

1. INTRODUCTION

Since the social science field has latent traits that cannot be observed directly, researchers use indicators to identify these traits. When latent traits (concepts) are not clearly expressed hypothetically, researchers often develop a scale to measure them. When scales are developed to measure latent traits like success, attitude, interest, and belief, there are two common measurement theories; namely, the Classical Test Theory (CTT) and Item Response Theory (IRT). Since this research focuses on CTT, this paper only explains this theory and is limited to CTT. The CTT is used in numerous scale development studies due to its typical implementation in the software, easy-to-understand structure, suitability for social sciences, and relatively weak assumptions.

Moreover, CTT results are similar to and have high-level relationship with results obtained from different theories (ex. IRT) in certain situations (DeVellis, 2006; Fan, 1998). However, it is essential to note that there are also disadvantages, such as item and person statistics being

*CONTACT: Abdullah Faruk Kilic ✉ abdullahfarukkilic@gmail.com Adıyaman University, Faculty of Education, Department of Educational Sciences, Division of Educational Measurement and Evaluation, Adıyaman, Türkiye.

e-ISSN: 2148-7456 /© IJATE 2022

dependent on the test and sample (Kohli et al., 2015). Therefore, the sampling procedure that must be representative of the population when developing a scale becomes an important subject. Otherwise, item statistics (discrimination and difficulty) will fail to reflect the reality.

CTT assumes that each score contains the true and error scores related to the examined trait. The normal distribution of the error score is another assumption. Although CTT seems to focus on the items, it focuses on the entire test (DeVellis, 2006). When developing CTT-based scales, it is reasonable to apply item analysis before factor analysis (Kline, 2000) because item analysis can help decide the items to be kept in or removed from the scale (Green & Salkind, 2014). For item analysis, it is necessary to focus on exploratory statistics, item difficulty, and discrimination (Kline, 2005). The validity of test scores depends on the item validity in the test. Especially when the unidimensional structure is considered, a high-level relationship between item analysis and factor analysis is found (Kline, 2000). Therefore, it is reasonable to collect evidence towards the validity and reliability of the scores obtained from the scale after conducting item analysis. Item validity is investigated during item analysis and is frequently determined by item discrimination.

On the other hand, item discrimination is commonly investigated with discrimination index (D) and item-total correlation. The D index compares the lowest and highest performance groups in the test (Kaplan & Saccuzzo, 2018). Accordingly, the difference between the correct numbers of the upper and lower 25% (or 33%) groups is taken and divided by the number of individuals in a group (Brown, 1988; Metsämuuronen, 2020a). Cureton (1957) suggested using 27% for the upper and lower groups. 27% is a critical ratio that separates the tails from the mean in the standard normal distribution of errors. Item discrimination is also the strength of the relationship between an item in the test and other items. Therefore, it also measures the item's relationship with the true score (DeVellis, 2006). In other words, it is the relationship between one item and all items. Therefore, it is called item-total correlation. Item-total correlation is investigated with phi coefficient, tetrachoric, biserial, and point-biserial correlation coefficients for binary (1-0) scored items and Pearson product-moment correlation coefficient for polytomous scored items (ex. open-ended tests) (Kline, 2000). It can be seen that some applications calculate correlation after reducing the investigated item score from the total score. That application was named the corrected item-total correlation (Macdonald & Paunonen, 2002). Values obtained without corrected item-total correlation are biased (Kline, 2000) since correction is essential, especially when 5-6 items are in the test (Kline, 2005). The correlation will be higher than its actual value as item scores will be included in the total score with no correction.

In unidimensional structures, when the item-total correlations are positive and high, these items can distinguish low and high-level individuals from each other in terms of the trait measured by the item, which is the basis of item discrimination. Item-total correlation values show that the item discrimination varies between -1 and 1, like the Pearson product-moment correlation coefficients (Brown, 1988). A negative item discrimination value indicates inverse discrimination between individuals with low and high ability in terms of the measured trait. Negatively discrimination means that while individuals with a high trait have a low score on the item, individuals with a low trait have a high score. The increased discrimination of an item with a positive value indicates that individuals with low and high trait levels are effectively distinguished (Macdonald & Paunonen, 2002). There is a cut-off point for item discrimination. Most researchers state that item-total correlation must be at least .30 (Kline, 2000; Nunnally & Bernstein, 1994).

The related literature review shows relatively more common discrimination coefficients, examples of which include the D index, point-biserial correlation coefficient, biserial correlation coefficient, phi coefficient, tetrachoric correlation coefficient, and rank biserial correlation coefficient. There are less common discrimination indexes such as the B index, the

agreement statistic, Davis discrimination index, Flanagan's correlation coefficient, Flanagan's corrected correlation coefficient, and phi/(phi max) coefficient (Liu, 2008). In other words, researchers related to the discrimination coefficient have always been in a search for what the best discrimination coefficient is since there are currently more than 20 discrimination coefficients available in the literature. Although item discrimination has been investigated for a long time, the research on this subject is still ongoing.

Some studies compare item discrimination indexes or recommend a new index when the current literature is reviewed. For example, Bazaldua et al. (2017) stated that the literature has complicated results regarding item discrimination and compared point-biserial, biserial, and point-biserial with the item-rest score, phi coefficient for binary data which categorize using median value, discrimination index. The estimators showed different performances in the analysis by differentiating test length, item difficulty, item discrimination, and test score distribution. In another study, Liu (2008) compared the point-biserial and biserial correlation coefficients with the D coefficient calculated with different lower and upper group percentages (10%, 27%, 33%, and 50%). Item-factor correlations showed the closest result to the item-total correlation. In recent years, Metsämuuronen (2020a) conducted research in order to generalize the D index, a simple and robust coefficient. D index that gives consistent results even when there are outliers is generalized for items scored in more than two categories while e vector properties are used in generalization. In addition, Metsämuuronen (2020b) recommended Somers' D index as an alternative to item-total correlation and corrected item-total correlation. As a result of the simulation study, the researcher found that Somers' D index estimated values below the real value for items with four and more categories.

Even when multi-dimensional structures are found in CTT-based scale development studies, it is seen that the item-total correlation or corrected item-total correlation is examined when examining the item discrimination (Ak & Alpullu, 2020; Akyıldız, 2020; Çalışkan, 2020; Tarhan & Yıldırım, 2021). However, such analysis might lead to underestimates of item discrimination. Therefore, items that should be included in the scale might be removed from the scale. To avoid item removal, item-factor correlation or corrected item-factor correlation might be investigated (see also Green & Salkind, 2014). However, such an approach requires much effort and fails to provide information about the entire test. Our study built on this problem aims to provide an alternative approach to investigate item discrimination of scales developed or adapted based on the Classical Test Theory (CTT). We proposed a new item discrimination index for two-dimensional structures and tested it using the Monte Carlo simulation under the conditions of sample size, the magnitude of item discrimination, inter-factor correlation, and the number of categories. The newly developed item discrimination index can determine the discrimination of each item at one time by considering the scale's dimensionality. The inter-factor correlation can be considered with this newly proposed index, and a direct relationship can be established between the score for the entire test and items.

Our study contributes to the literature by eliminating the mentioned limitations regarding item discrimination and providing evidence for item discrimination by considering the dimensionality and inter-factor correlation in two-dimensional structures. Therefore, this study is considered necessary and aims to contribute to the literature by a) recommending a new item discrimination index for two-dimensional structures, b) investigating the recommended item discrimination index under numerous simulation conditions, and c) the new recommended discrimination index can be used in scales development studies. The detailed information regarding this index is provided as follows:

1.1. New Index

A vector length in analytic geometry is used by considering the inter-factor correlation to develop a two-dimensional item discrimination index. The item discrimination values

calculated for each dimension of an item create a vector in the space. Let us consider a two-dimensional example: In a two-dimensional structure, an item's correlation with the first and second dimensions is expressed by two values, D_1 and D_2. These points can be represented as ordered pairs in two-dimensional Euclidean space, which is presented in Figure 1.

Figure 1. *D_1 and D_2 points on the plane.*

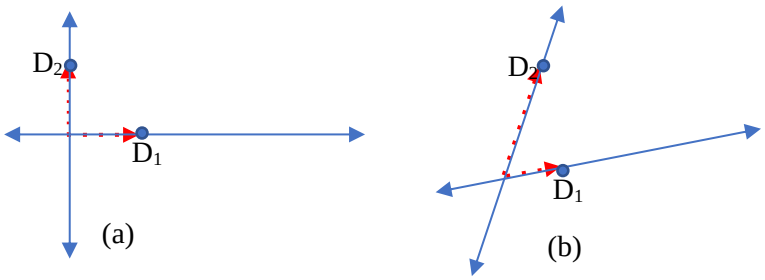


Figure 1-(a) shows that the x and y axes represent vertical two dimensions. D_1 and D_2 points represent the discrimination of an item in each dimension. D_1 and D_2 points can be represented as ordered pairs (a,0) and (0,b). At the same time, these points indicate a vector on a plane. Similarly, D_1 and D_2 points in Figure 1-(b) are points on the affine coordinate system. The affinity of the axes indicates a correlation between the dimensions. The correlation between the dimensions equals the cosine of the angle between these two vectors (Gorsuch, 1974). In this case, the product of these vectors is found to learn about the discrimination on both dimensions. The starting point of this study is this idea. The parallelogram method is applied to find out the product of these points, and the product vector is found as equation 1:

$$\overrightarrow{V_b^2} = a^2 + b^2 + 2abcos\theta \quad (1)$$

(Lange, 2009). Here, a represents the x-axis value, b represents the y-axis value, and θ represents the angle between the x and y axes. Since the axes will have a 90° angle when they are perpendicular, cos(90°) = 0 will give the resultant vector as $\overrightarrow{V_B} = \sqrt{a^2 + b^2}$. However, when the axis is affine, the coordinates on these affine systems are first transformed into the rectangular coordinate system. The product vector is calculated as in the perpendicular coordinate system. The transformation matrix in equation 2 is used for this transformation (Deakin, 1998).

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (2)$$

Accordingly, X' and Y' values correspond to discrimination in the affine coordinate system, while X and Y values are the correspondence in the rectangular coordinate system. θ is the angle between the two axes. When the equation system in Equation 2 is solved, X and Y values are obtained. Since X' and Y' values correlate with each item's dimension for two-dimensional structures, these are known as numerical values. θ value can be obtained from the correlation between two dimensions. Since the correlation (r_{xy}) between two dimensions is cos(θ) (Gorsuch, 1974), which will be $\arccos(r_{xy}) = \theta$, the value obtained here can be used for calculating $sin(\theta)$. Thus, two unknown values in the equation system will be X and Y. If this equation system is solved:

$$X' = cos\theta.X + sin\theta.Y \quad (3)$$

$$Y' = -sin\theta.X + cos\theta.Y \quad (4)$$

will be obtained. Here, if we multiply equation three to $(-cos\theta)$ and equation four to $(sin\theta)$, we obtain X and Y variables:

$$-cos\theta.X' = -\cos^2\theta\,.X - cos\theta.sin\theta.Y \quad 5$$

$$sin\theta.Y' = -\sin^2 X + sin\theta.cos\theta.Y \quad 6$$

equations are obtained. If each side of the Equations 5 and 6 are summed:

$$sin\theta.Y' - cos\theta.X' = (\sin^2\theta + \cos^2\theta).X \quad 7$$

$$X = sin\theta.Y' - cos\theta X' \quad 8$$

equations are obtained. Thus, the X variable is found. X variable can be written in Equation 3, and similar operations are followed for the Y variable:

$$Y = -sin\theta.X' + cos\theta.Y' \quad 9$$

by writing the X and Y variables obtained from here to Equation 1 a and b variables, a two-dimensional discrimination index is obtained.

1.2. An Example of a New Index

Let us assume that an item's discrimination index for the first dimension (correlation) is .50, and the discrimination index for the second dimension (correlation) is .20 on a two-dimensional scale, then the inter-factor correlation is .30. Let us calculate the two-dimensional discrimination coefficient of an item obtained from a two-dimensional scale: Here, $X' = .50$ and $Y' = .20$ because the X' and Y' values in the new discrimination index are the correlation of the item for two dimensions. Since the correlation between the two dimensions is given as .30, we have $cos\theta = .30 \Rightarrow arccos\,(0.30) = \theta$. Here, $\theta = 72.54^o$ is obtained. When these values are written to Equations 8 and 9:

$$X = \sin(72.54^o)\,.\,0.20 - \cos(72.54^o)\,.\,0.50 \quad 10$$

$$Y = -\sin(72.54^o)\,.\,0.50 + \cos(72.54^o)\,.\,0.20 \quad 11$$

equations are obtained. X and Y values are obtained as 0.0407 and -0.4169, respectively. When X and Y values are written to Equation 1 respectively as a and b and written to $cos\theta = .30$:

$$\xi = \sqrt{0.0407^2 + (-0.4169)^2 + 2.(0.0407).(-0.4169).0.30} \quad 12$$

The equation is obtained. ξ results as 0.4065 when equation 12 is completed. Accordingly, for a two-dimensional scale, the correlation of an item with the first and second dimensions is .50 and .20, respectively. The discrimination for both dimensions is obtained as .41 when two dimensions are considered together.

2. METHOD

This study investigated a new item discrimination index for two-dimensional structures in a Monte Carlo simulation. In Monte Carlo simulation studies, the data are generated to fit the desired distribution properties (Bandalos & Leite, 2013) and analyzed in line with the purpose of the study.

2.1. Simulation Conditions

In this study, the sample size (200, 500, and 1000), the magnitude of item discrimination (.30, .50, and .70), inter-factor correlation (.00, .30, and .50), and the number of categories (2, 3, 5 and 7) were the simulation conditions and the fixed simulation condition was two-dimensional structures (see the further details in the data analysis section).

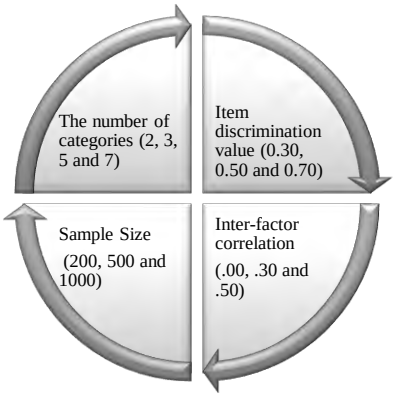
200, 500, and 1000 conditions were determined for the sample size. Kılıç and Koyuncu (2017) reviewed the scale adaptation studies in Turkey and found that more than half of the studies had samples between 100 and 349, and the average was 244. On the other hand, another study investigating the scale development studies in Turkey (Koyuncu & Kılıç, 2019) reported that more than half of the studies investigated included 300 or more individuals. Goretzko et al. (2021) systematically reviewed scale development studies and stated that more than half of the studies had a 400 or higher sample size. For this reason, considering the item discrimination index mainly reported for scale development and adaptation studies, the sample size was selected as 200 and 500. The 1000 sample size condition was included in this study to investigate the effects of increased sample size on the results.

.30, .50, and .70 conditions were determined for the magnitude of item discrimination. Since item discrimination between the .30-.39 range suggest that the item can be directly included in the scale/test (Crocker & Algina, 2008), the .30 condition was added to the study. On the other hand, since the item discrimination was desired to be .40 and above, the .50 and .70 conditions were added to the research as the conditions where the item discrimination was medium and high, respectively.

.00, .30 and .50 conditions were investigated for inter-factor correlation. .00 inter-factor correlation suggests no relationship between the dimensions, i.e., the dimensions are perpendicular. The .00 inter-factor correlation condition was added since item-total correlation was investigated while the item discrimination was calculated. Thus, it was aimed to examine the results that would emerge when the total score is taken in a situation where the total score should not be taken. On the other hand, the inter-factor correlation is generally reported and investigated as .30 in empirical (Li, 2016) and simulation studies (Cho et al., 2009; Curran et al., 1996; Flora & Curran, 2004; Foldnes & Grønneberg, 2017). Therefore, this simulation condition was added to the study. .70 inter-factor correlation condition was added due to high correlation between the dimensions in order to investigate the item-total score correlation results when getting a total score would cause no problems.

The number of categories of variables was manipulated as 2, 3, 5, and 7 in this study. The scale items are often Likert-type, and Likert-type items are generally scored as five-point scores (Lozano et al., 2008). Therefore, five was added as the category number to the study. On the other hand, three category conditions were added to the study since 3-point scales were used for children. Two conditions were added since there might be achievement tests with multiple options, yes/no, or a control list. Lastly, seven category number condition was included in the study to investigate the effects of increased category number on the discrimination index. Figure 2 briefly shows the simulation conditions.

Figure 2. *Simulation conditions.*

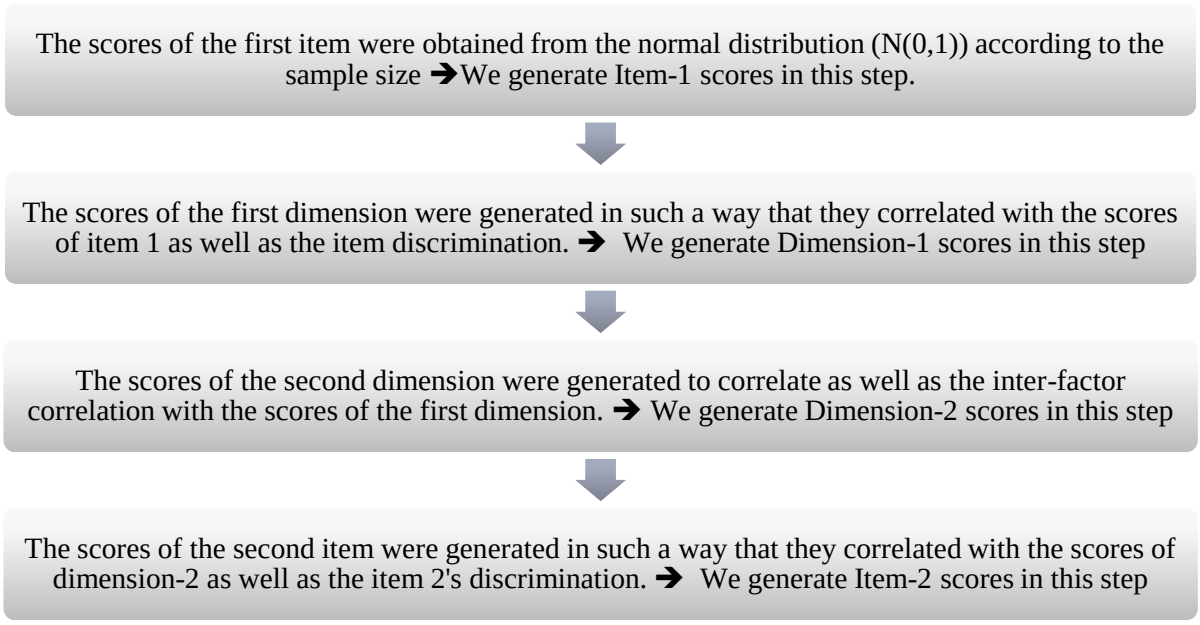


This simulation study was carried out with fully cross design. As seen in Figure 2, a fully crossed design was applied, and the simulation was run for 4x3x3x3=108 conditions. 1000 replication was applied for each condition.

2.2. Data Analysis

We used R software (R Core Team, 2021) for data generation and came up with four variables: the first was item 1 scores, the second one was total score for dimension 1, the third one was total score for dimension 2, and last one was item 2 scores. The data generation process was given in Figure 3. Also, the data generation R codes were added (see Appendix 1).

Figure 3. *Data generation process.*



After data generation was performed, we added the scores of dimension-1 and dimension-2 to obtain the total scale score. We calculated item-total correlation using items (item-1 and item-2) scores and total scale scores. Thus, the correlations of the items with the scores obtained from the whole test were examined.

The R software's stats (R Core Team, 2021) package was used to calculate the proposed two-dimensional item discrimination index. The item-total and item-factor correlation were examined and the proposed two-dimensional discrimination index worked under simulation conditions was determined. Therefore, the graphics show the item-factor correlation, item-total correlation, and two-dimensional item discrimination index results. These results were used for a descriptive inference. Additionally, a one-way analysis of variance (ANOVA) was applied to investigate which conditions would have a more effect on item discrimination.

3. FINDINGS

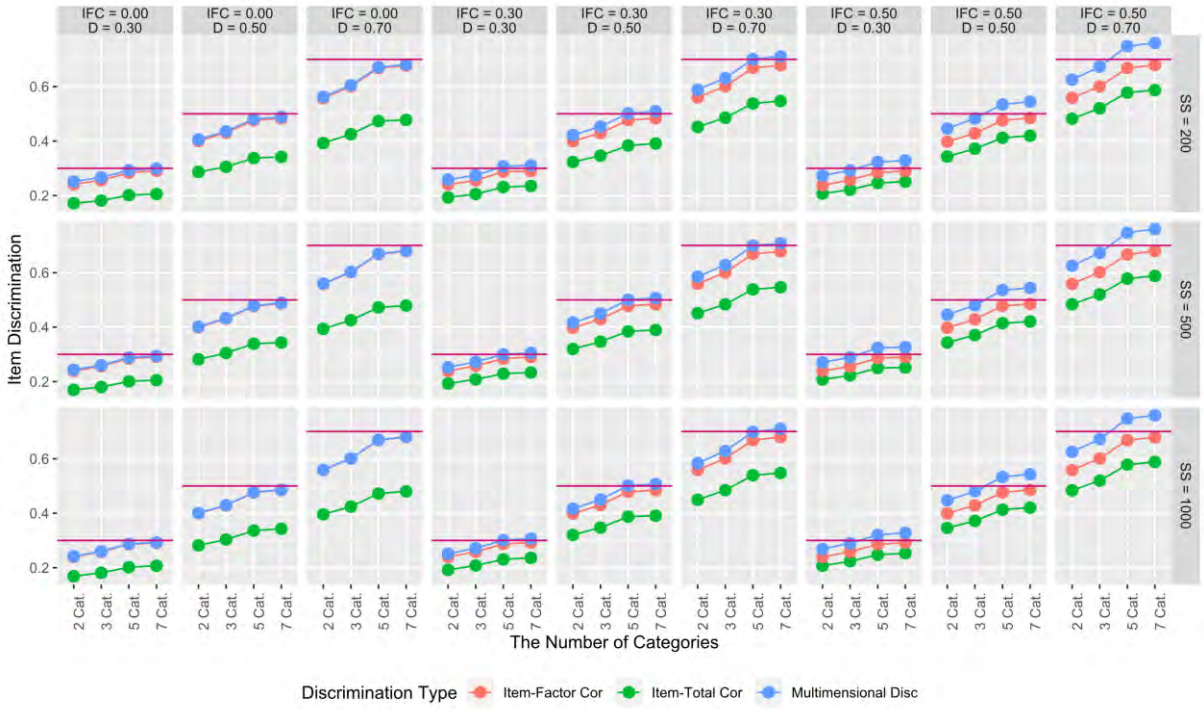
The average values obtained from item discrimination methods as a result of this study are given in Figure 4. Additionally, these values are given in the Appendix-2 for researchers who want to take a detailed look at these results. Figure 4 shows that the correlation between the dimensions is .00, i.e., when two dimensions are orthogonal, and the recommended two-dimensional discrimination index and the item-factor correlation revealed similar results. The calculated values are more accurate since the data became closer to continuous as the category number increased.

When the inter-factor correlation was .00, it could be stated that the item-total correlation was underestimated for all magnitude of item discrimination conditions. One reason is to examine

item-total correlation by taking the total score from two vertical dimensions. The simulation conditions with a .00 inter-factor correlation between the dimensions showed a better performance for the newly recommended method.

When the inter-factor correlation was .30, it is possible to say that the recommended item discrimination index had a higher value than that of the item-factor correlation. Since the correlation coefficients made a more reasonable estimation with the increased category number, the item discrimination indexes increased in 7-category items. However, the graphic shows that the corrected item-total score correlations fail to give results close to the actual values in any conditions.

Figure 4. *Discrimination indexes obtained from simulation conditions.*



When the conditions with a .50 inter-factor correlation and discrimination were investigated, it was observed that the two-dimensional discrimination index was overestimated. Although increasing the inter-factor correlation and magnitude of the item discrimination to .70 deviates the results of the recommended two-dimensional item discrimination index from the actual value, the value should be .70 and estimated as .76 at most. Accordingly, overestimation could be stated as approximately 9%.

The one-way analysis of variance conducted to investigate which simulation conditions affected the values obtained from the item discrimination methods revealed that the sample size had no significant effect on the item discrimination [$F_{(2,312)}$=.04, p=.97]. There is a significant difference between category number [$F_{(3,312)}$=189.21, p=.00], inter-factor correlation [$F_{(2,312)}$=70.33, p=.00], magnitude of item discrimination [$F_{(2,312)}$=4906.54, p=.00], and item discrimination methods [$F_{(2,312)}$=668.99, p=.00]. When the effect size was investigated, the eta-square value was found to be .97 for the magnitude of item discrimination conditions, .81 for item discrimination method, .65 for category number, and .31 for inter-factor correlation. Accordingly, the most impactful factor on item discrimination estimations was the magnitude of item discrimination. According to Green and Salkind (2014), the eta-square value of .14 shows a high impact size. Based on this, it could be stated that the eta-square values obtained for the magnitude of item discrimination, item discrimination method, category number, and inter-factor correlation had a significantly high impact.

4. DISCUSSION and CONCLUSION

A new item discrimination index was obtained for two-dimensional structures in the current study, which was carried out based on the discrepancies in examining item-total score correlations for item discrimination in multi-dimensional constructs based on a test score or factor score. After numerous investigations on category number, sample size, the magnitude of item discrimination, and inter-factor correlation based on Monte Carlo simulation, the newly obtained item discrimination index can be used for two-dimensional structures. This study shows a significant difference between item-total correlation, item-factor correlation, and recommended item discrimination for two-dimensional structures. This finding matches the results of Bazaldua et al. (2017) as they failed to find similarities between item discrimination methods when multiple item discrimination methods were compared. Such results support the hypothesis stated in the problem situation of the current research.

Item-factor correlations showed similar results with the newly recommended index, especially when the correlation between the factors was extremely low (correlation was taken as .00 to exemplify this condition in the current study). It can be seen that the recommended item discrimination index for two-dimensional structures showed adequate performance when the fact that item-factor correlations should be investigated for two-dimensional structures. Moreover, the recommended two-dimensional discrimination index could be used when the correlation between dimensions was extremely low. Also, the results in item-factor correlations provide ideas about the factor, not the entire test. Considering that item-total correlations underestimate the discrimination, it is beneficial to use the newly item discrimination index for two-dimensional structures that can be calculated at once.

When the inter-factor correlation increased to .30, although the item-total correlation was closer to the actual value of the item discrimination, the value was deficient. This situation may cause researchers to be mistaken when making decisions about items. Item-factor correlations of two-dimensional structures and newly item discrimination index revealed similar values. Although the item-factor correlations were highly close, these revealed slightly lower results than the actual values. When the factor correlation was .00 or .30, the item-total and item-factor score correlations showed differences. However, contrary to this finding, Liu (2008) stated that the item-total and item-factor correlations had similar results. It is believed that the difference between our specific study and Liu's (2008) study was due to mixed-format test usage.

As the inter-factor correlation increased to .50 and the discrimination value to .70, the highest value was obtained for two-dimensional structures in the new item discrimination index. When inter-factor correlation was .50 and item discrimination was .70, the behaviors of the item discrimination methods differentiated more. When the inter-factor correlation was .50, the item-total correlations were underestimated; the item-factor correlations were estimated close to the actual value, and the new two-dimensional item discrimination index was overestimated. The overestimation percentage for the new two-dimensional discrimination index was 9 at most.

The sample size was not found as a significant independent variable to impact the estimation of item discrimination. One of the reasons might be that the smallest sample size was 200. In addition, the magnitude of item discrimination and inter-factor correlations were found as significant independent variables. Therefore, inter-factor correlation and magnitude of item discrimination should be considered by researchers when item discrimination is investigated. It is important to note that the item discrimination index for two-dimensions might be slightly overestimated when the inter-factor correlation is high (approximately .70).

There is another important finding in this study. The methods can identify the item discrimination more accurately as the category number increases. However, Metsämuuronen (2020b) recommended that Somers' D coefficient estimates two-category data better. The difference between the new item discrimination index in the current study and

Metsämuuronen's (2020b) study is due to the different mathematical basis. The recommended index in this study is in line with corrected item-total and item-factor correlations. In addition, Metsämuuronen (2020a) generalized the D index for items scored in more than two categories by using vectors in the study to generalize the D index. Similarly, calculations of the discrimination index for two-dimensional structures in the current study were based on vectors.

The item discriminations were underestimated when the category number was low. However, the literature does not show different cut-off points for item-total correlations according to category number. Although there is no rule of thumb, the cut-off point for the new item discrimination index can be determined as .30 when data are in 5 and 7 categories. Considering that the data have 2 or 3 categories, it is rational to accept the new item discrimination index up to .20. Accordingly, different cut-off points can be determined for different categories and discrimination indexes by conducting simulation studies in future studies.

This study has certain limitations. Since the recommended discrimination index is newly developed, we investigate it for only two-dimensional structures. Future studies can focus on three or more dimensional structures. Moreover, the item discrimination index for two-dimensional structures might be revised based on the studies with 3, 4, 5, or higher dimensions and added to open-source software (Python, R, etc.). This study has not covered items with cross-loading. In future studies, the performance of the developed item discrimination index can be examined in cases where items have cross-loading.

The item discrimination index for two-dimensional structures revealed as a result of this study can be recommended only for two-dimensional structures. We named the recommended item discrimination index as ξ coefficient. Therefore, researchers using the recommended index for two-dimensional structure could show the index as a ξ coefficient.

Acknowledgments

This study was presented as an oral presentation at the International Congresses on Education (ERPA) in 2021.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Abdullah Faruk Kilic: Investigation, methodology, developing process, visualization, software, formal analysis, and writing-original draft. **Ibrahim Uysal**: Developing process, methodology, resources, validation, and writing-original draft.

Orcid

Abdullah Faruk Kilic https://orcid.org/0000-0003-3129-1763
Ibrahim Uysal https://orcid.org/0000-0002-6767-0362

REFERENCES

Ak, M.O., & Alpullu, A. (2020). Alpak akış ölçeği geliştirme ve Doğu Batı üniversitelerinin karşılaştırılması [Alpak flow scale development and comparison of east west universities]. *E-Journal of New World Sciences Academy, 15*(1), 1-16. https://doi.org/10.12739/NWSA.2019.14.4.2B0122

Akyıldız, S. (2020). Eğitim programı okuryazarlığı kavramının kavramsal yönden analizi: Bir ölçek geliştirme çalışması [A conceptual analysis of curriculum literacy concept: A study of scale development]. *Electronic Journal of Social Sciences, 19*(73), 315–332. https://doi.org/10.17755/esosder.554205

Bandalos, D.L., & Leite, W. (2013). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.). Information Age.

Bazaldua, D.A.L., Lee, Y.-S., Keller, B., & Fellers, L. (2017). Assessing the performance of classical test theory item discrimination estimators in Monte Carlo simulations. *Asia Pacific Education Review, 18*, 585–598. https://doi.org/10.1007/s12564-017-9507-4

Brown, J.D. (1988). Tailored cloze: Improved with classical item analysis techniques. *Language Testing, 5*(1), 19–31. https://doi.org/10.1177/026553228800500102

Cho, S.-J., Li, F., & Bandalos, D.L. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748–759. https://doi.org/10.1177/0013164409332229

Crocker, L., & Algina, J. (2008). *Introduction of classical and modern test theory*. Cengage Learning.

Cureton, E.E. (1957). The upper and lower twenty-seven per cent rule. *Psychometrika, 22*, 293-296. https://doi.org/10.1007/BF02289130

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Çalışkan, A. (2020). Kriz yönetimi: Bir ölçek geliştirme çalışması [Crisis management: A scale development study]. *Journal of Turkish Social Sciences Research, 5*(2), 106–120.

Deakin, R. (1998). 3-D coordinate transformations. *Surveying and Land Information Systems, 58*(4), 223–234.

DeVellis, R.F. (2006). Classical test theory. *Medical Care, 44*(11), 50-59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58*(3), 357–381. https://doi.org/10.1177/0013164498058003001

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Foldnes, N., & Grønneberg, S. (2017). The asymptotic covariance matrix and its use in simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(6), 881–896. https://doi.org/10.1080/10705511.2017.1341320

Goretzko, D., Pham, T.T.H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, 40*(7), 3510–3521. https://doi.org/10.1007/s12144-019-00300-2

Gorsuch, R.L. (1974). *Factor analysis*. W. B. Saunders.

Green, S.B., & Salkind, N.J. (2014). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (7th ed.). Pearson Education.

Kaplan, R.M., & Saccuzzo, D.P. (2018). *Psychological testing: Principles, applications, and issues*. Cengage Learning.

Kılıç, A.F., & Koyuncu, İ. (2017). Ölçek uyarlama çalışmalarının yapı geçerliği açısından incelenmesi [Examination of scale adaptation studies in terms of construct validity]. In Ö. Demirel & S. Dinçer (Ed.), *Küreselleşen dünyada eğitim* [Education in the globalized world] (pp. 1202–1205). Pegem.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). Routledge.

Kline, T.J.B. (2005). *Psychological testing: A practical approach to design and evaluation* (3rd ed.). Sage.

Kohli, N., Koran, J., & Henn, L. (2015). Relationships among classical test theory and item response theory frameworks via factor analytic models. *Educational and Psychological Measurement, 75*(3), 389–405. https://doi.org/10.1177/0013164414559071

Koyuncu, İ., & Kılıç, A.F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Education and Science, 44*(198), 361-388. https://doi.org/10.15390/EB.2019.7665

Lange, M. (2009). A tale of two vectors. *Dialectica, 63*(4), 397-431. https://doi.org/10.1111/j.1746-8361.2009.01207.x

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Liu, F. (2008). Comparison of several popular discrimination indices based on different criteria and their application in item analysis [Master of Arts, University of Georgia]. http://getd.libs.uga.edu/pdfs/liu_fu_200808_ma.pdf

Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*(2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Macdonald, P., & Paunonen, S.V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62*(6), 921-943. https://doi.org/10.1177/0013164402238082

Metsämuuronen, J. (2020a). Generalized discrimination index. *International Journal of Educational Methodology, 6*(2), 237-257. https://doi.org/10.12973/ijem.6.2.237

Metsämuuronen, J. (2020b). Somers' D as an alternative for the item–test and item-rest correlation coefficients in the educational measurement settings. *International Journal of Educational Methodology, 6*(1), 207-221. https://doi.org/10.12973/ijem.6.1.207

Nunnally, J.C., & Bernstein, I.H. (1994). *Pschometric theory* (3rd ed.). McGraw Hill.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. https://www.r-project.org/

Tarhan, M., & Yıldırım, A. (2021). Bir ölçek geliştirme çalışması: Hemşirelikte geçiş şoku ölçeği [A scale development study: Nursing Transition Shock Scale]. University of *Health Sciences Journal of Nursing, 3*(1), 7-14. https://doi.org/10.48071/sbuhemsirelik.818123

APPENDIX

Appendix 1. *R Codes for data generation.*

```
generate_data <-   function(seed, discrimination, interfactor_cor, sample_size) {
  #Set the seed and generate the parameters
  set.seed(seed)
  i_1 <- rnorm(sample_size, 0, 1)
  t_1 <- rnorm(sample_size, discrimination*i_1, sqrt(1-discrimination^2))
  t_2 <- rnorm(sample_size, interfactor_cor*t_1, sqrt(1-interfactor_cor^2))
  i_2 <- rnorm(sample_size, discrimination*t_2, sqrt(1-discrimination^2))
  tidyr::tibble(i_1, i_2, t_1, t_2, scale_score = t_1 + t_2)
}
```

Appendix 2. *Discrimination indexes values obtained from simulation conditions.*

The Number of Categories	Type of Item Discrimination	Sample Size: 200									Sample Size: 500									Sample Size: 1000								
		Inter-factor Correlation: .00			.30			.50			.00			.30			.50			.00			.30			.50		
		Item Discrimination: .30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70	.30	.50	.70
2	IFC	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56	.24	.40	.56
	ITC	.17	.29	.39	.19	.32	.45	.21	.34	.48	.17	.28	.39	.19	.32	.45	.21	.34	.48	.17	.28	.40	.19	.32	.45	.21	.35	.48
	MD	.25	.41	.56	.26	.42	.59	.27	.45	.63	.24	.40	.56	.25	.42	.59	.27	.45	.63	.24	.40	.56	.25	.42	.58	.27	.45	.63
3	IFC	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60	.26	.43	.60
	ITC	.18	.31	.43	.21	.35	.49	.22	.37	.52	.18	.31	.43	.21	.35	.48	.22	.37	.52	.18	.30	.42	.21	.35	.48	.22	.37	.52
	MD	.27	.44	.61	.27	.45	.63	.29	.48	.67	.26	.43	.60	.27	.45	.63	.29	.48	.67	.26	.43	.60	.27	.45	.63	.29	.48	.67
5	IFC	.28	.48	.67	.29	.48	.67	.28	.48	.67	.28	.48	.67	.28	.48	.67	.29	.48	.67	.29	.48	.67	.29	.48	.67	.29	.48	.67
	ITC	.20	.34	.47	.23	.38	.54	.25	.41	.58	.20	.34	.47	.23	.38	.54	.25	.41	.58	.20	.34	.47	.23	.39	.54	.25	.41	.58
	MD	.29	.48	.67	.31	.50	.70	.32	.53	.75	.29	.48	.67	.30	.50	.70	.32	.54	.75	.29	.48	.67	.30	.50	.70	.32	.53	.75
7	IFC	.29	.48	.68	.29	.48	.68	.29	.48	.68	.29	.49	.68	.29	.48	.68	.29	.49	.68	.29	.49	.68	.29	.48	.68	.29	.49	.68
	ITC	.21	.34	.48	.24	.39	.55	.25	.42	.59	.21	.34	.48	.23	.39	.55	.25	.42	.59	.21	.34	.48	.24	.39	.55	.25	.42	.59
	MD	.30	.49	.68	.31	.51	.71	.33	.54	.76	.29	.49	.68	.31	.51	.71	.33	.54	.76	.29	.49	.68	.31	.51	.71	.33	.54	.76

IFC: Item-Factor Correlation, ITC: Item-Total Correlation, MD: Multidimensional Discrimination