International Journal of Assessment Tools in Education
2022, Vol. 9, No. 4, 808–830

https://doi.org/10.21449/ijate.1059200

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate **Research Article**

The comparison of the dimensionality results provided by the automated item selection procedure and DETECT analysis

Ezgi Mor[1,*], **Seval Kula-Kartal**[2]

[1]Kastamonu University, Faculty of Education, Department of Educational Sciences, Türkiye
[2]Pamukkale University, Faculty of Education, Department of Educational Sciences, Türkiye

ARTICLE HISTORY

Received: Jan. 17, 2022
Revised: Aug. 23, 2022
Accepted: Dec. 13, 2022

Keywords:
Dimensionality,
Mokken scaling DETECT,
Automated item selection procedure,
Non-parametric item response theory.

Abstract: The dimensionality is one of the most investigated concepts in the psychological assessment, and there are many ways to determine the dimensionality of a measured construct. The Automated Item Selection Procedure (AISP) and the DETECT are non-parametric methods aiming to determine the factorial structure of a data set. In the current study, dimensionality results provided by the two methods were compared based on the original factorial structure defined by the scale developers. For the comparison of the two methods, the data was obtained by implementing a scale measuring academic dishonesty levels of bachelor students. The scale was conducted on junior students studying at a public and a private university. The dataset was analyzed by using the AISP and DETECT analyses. The "mokken" and "sirt" packages on the R program were utilized for the AISP and DETECT analyses, respectively. The similarities and differences between the findings provided by the methods were analyzed depending on the original factor structure of the scale verified by the scale developers.

1. INTRODUCTION

In social sciences, the traits mostly studied are complex, and have an abstract structure that is generally composed of several different components. Researchers frequently employ the exploratory techniques to explore the assessed constructs, and they endeavor to find out the relationships between the constructs and theories. Discovering these associations provides evidence to confirm or invalidate theoretical propositions (Antino et al., 2018). The researchers analyze structures of the latent constructs in detail by employing different dimensionality approaches. Therefore, the investigation of the dimensionality analyses has been an essential part of examining a psychological construct.

The dimensionality has been defined as the minimum number of latent traits which is required to describe the statistical dependency in the data (Zhang & Stout, 1999). If the structure of the data can be explained by only one latent trait, then the dimensionality turns into the unidimensionality. Unidimensionality means that a set of items composing a scale measure only

*CONTACT: Ezgi Mor ezgimor@gmail.com Kastamonu University, Faculty of Education, Department of Educational Sciences, Turkiye

e-ISSN: 2148-7456 /© **IJATE** 2022

one psychological trait (Hattie, 1985). It refers to the existence of only one underlying dimension accounting for the variation in examinee responses. The items of a unidimensional scale purport to measure a single attribute (Sick, 2010). Hence the interpretation of the total score becomes easier and more meaningful. However, unidimensionality may not be valid for each data set. Most of the latent traits targeted by the measurement tools tend to be multidimensional due to the complex nature of psychological constructs (Hemker et al., 1995). Since the targeted traits generally have complex structures, it is very likely to observe multidimensionality in a given dataset.

Multidimensionality in a dataset might be introduced in several ways, because there are many factors affecting respondents' performances on a test apart from the assessed latent trait. These factors might be the personal ones, such as the level of motivation, anxiety, and fatigue etc., or testing factors such as local dependence of the items. However, if the test assesses one dominant dimension, the mentioned factors affect the respondents' performance as minor factors. The dominant dimension reflects the targeted trait with the test, and it determines the success levels of the respondents on the test, hence the test is accepted as unidimensional (Stout, 1999). Considering the complex structure of the dimensionality and unidimensionality issues, it is an undeniable fact that intensive analyses should be employed by the researchers to determine the dimensionality of the traits correctly.

Messick (1975) stated that to assess the meaningfulness of the inferences made from test scores, test developers should confirm what the test score itself actually exhibits. Hence, to make meaningful, appropriate, and useful inferences from the test scores, the construct validity of the scores should be examined meticulously (Kane, 2006; Lissitz, 2009; Sireci, 2009; Zumbo, 2009). Investigation of dimensionality of the measured trait or the structure of the phenomenon is an inevitable part of the construct validity (Slocum-Gori & Zumbo, 2011). Based on discussions on dimensionality, evaluation of dimensionality is a required stage in gathering evidence to support the validity of inferences made from total scores (Yu et al., 2007).

Many methods have been proposed by researchers to investigate the dimensionality of a dataset. For the last 30 years, the two notable reviews of methods and indices of the unidimensionality have been conducted. One of the first studies was conducted by Hattie (1985), in this research the researcher reviewed numerous approaches, and revealed weak sides of these approaches. Tate (2003) expanded the findings of Hattie’s (1985) study and included a review of methods and indices applied to discrete variables. In addition, the researcher stated that the most of the available methods perform effectively “within the assumptions”. It can be stated that the parametric dimensionality techniques such as the factor analytic methods have strict assumptions to be met to provide accurate results concerning dimensional structure of a dataset. Hence, to assess the dimensionality of the data, there has been an increasing interest in the use of nonparametric techniques and there is increasing number of studies comparing these techniques. To investigate the internal structure of the scales composed of dichotomous items, several researchers have suggested using the Mokken scale analysis (MSA) (Hemker, Sijtsma, and Molenaar 1995; Mokken 1971; van der Eijk and Rose 2015; van Schurr 2003). In addition to these researches proposing MSA, there are several research stuides in which the parametric and nonparametric techniques are compared and the advantages of the drawbacks are analyzed (Finch 2010, 2011; Kuijpers, van der Ark, and Croon 2013; van Abswoude, van der Ark, and Sijtsma 2004; Wismeijer et al. 2008). Wismeijer et al. (2008) compared the results of PCA and MSA with the real data set gathered by Self-Concealment Scale. They proposed the MSA as a complementary tool to PCA to determine the dimensionality of a data set. The scalability coefficients produced by the MSA and the different cutoff values, c values, were cited as the advantages of the MSA over the PCA. They recommended the usage of the MSA in addition to the PCA especially in the decision of the items’ retaining or discarding from the scale.

One of the latest researches conducted by Antino et al. (2018) compared MSA with factorial analysis models under conditions of multidimensionality. The researchers compared the nonparametric techniques MSA, item factor analysis (IFA) and Normal Ogive Harmonic Analysis Robust Method (NOHARM). The results of the study proved that MSA should be used as a tool to allocate the items after the unidimensionality is ensured with other methods. The MSA results indicated that items from different but correlated latent dimensions may be grouped as in the same dimension. Eijk and Rose (2015) also stated that the application of MSA is recommended only when the latent structure is refined well.

The popularity of the nonparametric methods is not surprising because they are generally based on less restrictive assumptions than parametric methods. In addition, these methods allow researchers to analyze the dimensionality of datasets obtained from smaller samples (Stout et al., 2002). In line with these advantages of the nonparametric methods, many studies have examined alternative nonparametric methods to analyze the dimensionality of a dataset. Some methods suggested in the related studies have been widely accepted and used by researchers. They are the DIMTEST (Stout, 1987; 1990), the DETECT (Kim, 1994; Zhang & Stout, 1999), and the Hierarchical Cluster Analysis with Proximity Matrix (Roussos et al., 1998). These three methods are all nonparametric statistical analyses. One of the more recently proposed nonparametric methods to analyze the dimensionality of a dataset is the Automated Item Selection Procedure (AISP) of the Nonparametric Item Response Theory (NIRT) approach. The AISP is also known as Mokken Scale Analysis (MSA) (Sijtsma et al., 2011).

The comparative research studies that investigate the performances of different nonparametric dimensionality assessment methods were mostly conducted on simulated data sets. Several studies reported that the performance of the AISP is inferior to the alternative nonparametric techniques in demonstrating the correct dimensionality of the data set (Mroch & Bolt, 2006; van Abswoude et al., 2004). Specifically, it was found that if the components of the latent constructs are correlated, the MSA may produce more erroneous results, and item may load on more than one dimension at the same time (Mroch & Bolt, 2006; van Abswoude et al., 2004). It should be noted that these results were obtained from the simulated datasets, and despite the stated drawbacks, the MSA and AISP methods have been kept using in determination of dimensionality of the assessed traits (Emons et al., 2010; Koster et al., 2009; Meijer et al., 2011; Ordon˜ez et al., 2009; Roorda et al., 2011; Sousa et al., 2010). In Stout et al.'s study (1996), the results obtained from the AISP were compared with the results obtained from the DIMTEST. The researchers found that the AISP has the advantage that it agrees with measurement practice in personality measurement to form facet scales. In addition, it has been still recommended to be used in the dimensionality analyses (Sijtsma et al., 2011). Therefore, the researchers have concluded that there is still a need to investigate the performance of the AISP, especially on empirical data sets. Therefore, the current study aimed to analyze dimensionality results provided by the two nonparametric techniques, the AISP and DETECT for a real dataset. More detailed information for the AISP and DETECT analyses were given in the following.

1.1. The Automated Item Selection Procedure (AISP)

The AISP is a technique that provides a way to investigate the dimensionality assumption in the context of the NIRT approach. This procedure is primarily based on the inter-item covariances, and the strength of the relationship between items and the assessed trait(s). This procedure reveals homogenous subscales of a scale based on the item covariances and item discrimination indexes called as scalability coefficient in the NIRT. While determining item discriminations, it also allows discarding the low-quality items out of the analysis. It results in clustering of test items with reasonable discrimination power that measure the same latent trait, and it composes a unidimensional scale from a large item pool. From this point of view, it can be used to analyze the dimensionality of scales and investigate the psychometric properties of scales (Meijer & Baneke, 2004; Sijtsma & Molenaar, 2002).

The AISP takes the raw dataset as input and reveals the dimensionality structure of the dataset. While doing this, the AISP uses the scalability coefficients of H (Loevinger, 1948; Molenaar, 1991). Scalability coefficients have crucial importance for MSA that works by pursuing unidimensional scales based on the Loevinger's definition of homogeneity and H coefficients. Scalability coefficients are related with homogeneity which is denoting the unidimensionality of a measure and MSA employs these coefficients to compose unidimensional scales. The H coefficients are defined at three levels: the item (Hi), item pair (Hij), and the whole scale level (H). These coefficients can be expressed as ratios of observed covariance and maximum possible covariance (Meijer, et al., 2015). The first step of MSA is testing the hypothesis about the scalability coefficients. These hypotheses are 1) For each item pairs, item pair scalability coefficients are calculated, and these coefficients show the covariance between two ordered variables. This index expresses the degree to which two items may belong to the same dimension. 2) Like item pair scalability coefficients, item level scalability coefficients are estimated that articulating how much an item is correlated to the sum score based on the remaining items. 3) The last hypothesis is based on the whole scale, as a complete set of the items, there is a test scalability coefficient showing the degree to which the total scores rank the test-taker on the assessed trait accurately. This index reaches a value of 1 when the scale is perfectly unidimensional (van Schuur, 2003). It has conventially been accepted to be higher than 0.30 (Mokken & Lewis, 1982).

Within the AISP, these coefficients are compared with a suitably chosen positive constant lower bound value, which is represented by the c. These coefficients are evaluated according to the lower bound value-constant (c) suggested by Mokken (1971, p.185). The c value is often accepted as 0.3, and items having Hi coefficients higher than 0.3 are included in the scale. For interpretation of all kinds of H values, the guidelines defined by Mokken (1971) are generally accepted. These guidelines are:

$.30 \leq H < .40$: items form a weak scale,

$.40 \leq H < .50$: items form a medium scale,

$.50 \leq H \leq 1.00$: items form a strong scale in terms of discrimination power.

The H coefficient of the scale is estimated from the Hi coefficients of the items. Therefore, power of a scale to discriminate among test-takers is dependent on whether scale items have high scalability coefficients or not. The power of the scale to measure the intended trait and provide an accurate ordering of individuals is determined based on some benchmark values. However, as Meijer et al. (2015) stated, there is no satisfactory level of studies explaining the meaning of these benchmarks. For that reason, the researchers have been advised to select different c values to control the quality of the scale.

There are also some problematic issues about the c values. In practice, higher values of scalability coefficients imply better item discrimination, the researchers may want to higher positive lower bound c. However, it doesn't always mean that high scalability coefficients will compose a discriminating unidimensional scale. In case of multiple latent variables models, the values of the Hij indexes may change according to two types of relationships. If the two items belong to the same latent dimension, the Hij index will show the impact of the factorial loading between each item and the common latent variable. In the second situation, if two items belong to the different latent dimensions, the Hij index will show the factorial loading of each item with its respective dimension, which is calculated as the multiplication of the correlation between two latent dimensions. This may cause a problem especially when the items are highly correlated with each other, and their discrimination indexes are high. They get higher Hij values as a product of correlations between each other, and even if they belong to the different latent dimension they may be grouped as in the same dimension. Hence based on AISP, the multidimensional scale may be erroneously accepted as a unidimensional scale (Antino et. al, 2018).

The AISP is a "bottom-up" procedure starting from selecting the pair of items of which a) the inter-item covariance, Hij, is higher than 0 significantly, and b) the Hij is the largest among the coefficients for all possible item pairs. Then, the third item is selected from the remaining items based on the levels of Hi coefficients. For the third item, (c) the Hi should be significantly higher than the 0, (d) it should be positively correlated with the first selected item-pair, and (e) the Hi coefficient should be higher than the selected benchmark for the scalability coefficients (c values). Thus, this process continues as long as items meeting specified conditions (c, d, e) are available. At the end of the process, the results might reveal more homogenous item clusters measuring different latent traits or latent trait composites (Meijer & Baneke, 2004). The interpretation of the clusters can be done based on the content of the items composing the same cluster. Lastly, a unidimensional scale is composed which provides a reasonable and reliable ranking of individuals on the latent trait by using their total scale scores (Sijtsma & Molenaar, 2002).

Suppose one wants to reach a scale with high reliability especially for a specified trait range. In that case, it is necessary to select highly discriminative items with item difficulties that span the desired range on the trait continuum. It might be very difficult to measure the whole trait continuum with the same level of precision; therefore, researchers may want to focus on one or more trait level. Sijtsma and Molenaar (2002) showed that items selected in the bottom-up procedure used in the AISP discriminate well across a wide range of item difficulties.

The other advantage of the AISP is that if multidimensionality is suspected in an empirical data set, well-chosen lower bound values will provide critical information about the dimensionality structure of the trait (Hemker et al., 1995). They suggested running the algorithm more than once with different lower bounds, c values, varying between 0.0 and 0.55. For a multidimensional structure, the AISP with varying lower bounds might result with the expected patterns such as: a) the most or all items belonging to one scale, b) items belonging to the two or more unidimensional scales, c) two or more scales including fewer items, and some items that need to be discarded from the procedure. Hemker et al. (1995) stated that the second step should be accepted as a result. As for unidimensional structure, the algorithm provides three sequential steps in case of the varying lower bounds. Firstly, most items are included in one scale; secondly, one smaller scale is detected with the increase in the lower bound. Lastly, one or several scales are determined, and some items are rejected. In this case, the result of the first step should be accepted as final. These findings revealed that the AISP may be used for unidimensional and multidimensional traits considering the different lower bounds for scalability coefficients. In addition, this feature of the AISP may provide a way to scale items that do not fit to any of the parametric IRT models (Reise & Waller, 2003). Hence, it can be concluded that using the AISP makes it possible to compose scales without conceding the content validity.

The AISP provides information for the psychometric qualities of items, therefore using the results of the AISP for building an item bank with already known psychometric properties is more suitable than utilizing the AISP in the context of construction of a scale based on the raw dataset. In addition, researchers are strongly advised to predict the dimensional structure of their item set based on the related theoretical foundation or the content of items. This makes easier to interpret the results of this procedure, and especially when the item set is not unidimensional, the findings can be put into better perspective (Sijtsma & Molenaar, 2002). Based on this suggestion, in the current study, a simulated data set was not generated, instead, a multidimensional scale whose psychometric qualities were already examined by the scale developers was preferred to evaluate performance of the AISP more efficiently.

1.2. The DETECT Analysis

The other method used to compare the results of the AISP is the DETECT technique. The DETECT provides information regarding the dimensionality of a dataset by enabling evidence for amount of multidimensionality. The main principles of the analysis are to specify the magnitude of dimensionality, test structure and the number of the dominant latent dimensions accounting for the inter-item covariances. It reveals whether an approximate simple structure underlies the item response data. The DETECT provides an index that is defined as the average of all signed conditional covariances calculated for item pairs. Suppose there is only one latent dimension influencing the item responses. In that case, the conditional covariances obtained from some item pairs will be positive while they will be negative for some item pairs. This will result with a low DETECT index since it is calculated based on the average of all signed covariances. However, if more than one dimension is underlying the test data, positive conditional covariances for the items within the same clusters, and negative conditional covariances for the items in distinct clusters will be explored. This will result with a higher DETECT index, which shows that the item response data departs from the unidimensionality and simple structure (Ackerman et al., 2003; Stout et al., 1996).

The DETECT aims at determining cluster partition providing the highest index. To reveal that partition, it calculates the index for different cluster partitions. The DETECT index is designed to be higher when calculated based on a cluster partition that is close to approximate simple structure. It uses different algorithms such as hierarchical cluster analysis to define cluster partition that produces the highest index. The partition giving the highest index determines the maximum value of the DETECT index. When this maximum value is equal or less than 0.10, it shows that one dominant dimension underlies the dataset. A maximum value between 0.10 and 0.50 indicates a weak amount of dimensionality; an index between 0.51 and 1.00 indicates a moderate amount of dimensionality. When the DETECT index is higher than 1.00, it can be accepted that strong amount of dimensionality exists in the data (Roussos & Özbek, 2006; Stout et al., 1996; Tate, 2003)

If the DETECT index reveals that the data differ from the (essential) unidimensionality, then, determining the dimensional structure gains importance. Another index, r, which is also estimated by the DETECT analysis provides information for the structure. This index is computed by dividing the maximum index by the sum of the absolute values of conditional covariances, which are calculated based on the cluster partition that gives the highest DETECT index. An r index between 0.80 and 1.00 indicates that the data is close to approximate simple structure, which means that test items form dimensionally homogenous clusters that are distinct from other clusters. The indexes produced by the DETECT provide answers to the three significant questions regarding the dimensionality of a dataset: Does the item response data hold (essential) the unidimensionality assumption? What is the amount of multidimensionality observed in the data? How many dominant dimensions account for the variation existed in the data? The analysis reveals the amount of multidimensionality exists in the data. Furthermore, if it is concluded that there is more than one dominant dimension accounting for item covariances, the DETECT provides a way to explore the dimensional structure and the number of dominant dimensions (Nandakumar, & Ackerman, 2004; Yu, & Nandakumar, 2001).

When the properties of the AISP and the DETECT methods are examined, it is clear that both techniques aim to discover the dimensionality of a dataset. Both techniques are nonparametric and require fewer assumptions than the parametric methods. The parametric techniques, such as the explanatory (EFA) and confirmatory factor analysis (CFA), are widely known and used by the researchers. However, despite the popularity of these methods, the factor analytic methods may sometimes perform inadequately, because they may confound the variation caused by item difficulty. As a result, the true number of latent factors is generally overestimated, hence, the findings may cause researchers to make erroneous inferences while interpreting individuals' total test scores (Stout et al., 1996). This situation is valid especially for dichotomous data. When test items are dichotomously scored, the Pearson matrix should be replaced with the tetrachoric matrix. However, the usage of this matrix for the factor analysis may not create common factors unless normality assumptions are met (Lord & Novick, 1968). In addition, if the sample size is less than 200, and item difficulties vary too much, the results of the tetrachoric matrix may not be dependable (Roznowski et al., 1994).

The parametric techniques may not always be suitable for analyzing a dataset's dimensionality of a dataset due to the difficulties in meeting the required assumptions. Furthermore, the parametric methods may result with the erroneous factorial solutions for the data if the researcher insists on using the parametric method although the data fail in meeting the necessary assumptions of the analysis. Therefore, it may be more accurate to utilize both the nonparametric and parametric methods to analyze the dimensionality of a dataset to lessen the possibility of obtaining erroneous results concerning the dimensional structure of the data. If findings obtained from the parametric and nonparametric methods are compatible, this will provide stronger evidence for the dimensional structure of the data. In the current study, the dimensional structure of a psychological trait, which was previously examined based on a parametric dimensionality technique (the exploratory or confirmatory factor analysis) will be determined based on the two nonparametric techniques: the AISP and DETECT procedures.

Theoretically, determining dimensional structure of a psychological trait is one of the most important steps of the test construction and analysis process. However, there is a very limited number of studies empirically investigating dimensionality of a dataset based on the nonparametric methods, especially the AISP procedure (Antino et al. 2018; Hemker et al. 1995; van Abswoude et al. 2004,). Therefore, it is envisaged that the present study will guide researchers to analyze the dimensionality of their data based on the nonparametric approach, which is expected to be great importance to researchers in educational and psychological measurement community and test developers in many fields. Since empirical studies examining the findings of dimensionality provided by nonparametric techniques are very rare, it is expected that findings of the study will contribute to the related empirical knowledge. Accordingly, the current study aims to analyze dimensionality of the dataset obtained from the implementation of the Academic Dishonesty Tendency Scale based on the AISP and DETECT methods. In addition, the CFA was carried out to validate the data gathered by the scale. Hence, the secondary purpose of the study is to compare the results provided by the nonparametric techniques with the results of the parametric one (confirmatory factor analysis) to examine whether the factorial solutions provided by the methods based on different approaches vary significantly. It is expected that revealing the differences and similarities among the dimensionality results provided by these techniques and providing detailed explanation and guidance on how to apply these techniques on the data and interpret the results of them will provide important information to the researchers interested in dimensionality analyzes.

2. METHOD

2.1. Research Model

This is a quantitative research study validating the factorial structure of a scale measuring the academic dishonesty of the undergraduate students based on the three methods, the CFA, AISP and DETECT. Considering the main goal of this study, it is suitable to define the study as a basic study.

2.2. Study Group

To gather the data of the study, undergraduate students of a public and a private university in Türkiye were included in the study group. It was not aimed to generalize the findings of the current study to population; therefore, instead of composing a random sample, convenient sampling was utilized based on the purposive sampling method. The scale aims to assess the academic dishonesty. The researchers thought that only the students who took the methods of scientific research course before may be aware of the concept of academic dishonesty. Therefore, the study group included junior students who had taken and succeeded the methods of scientific research course. The study group consisted of 212 junior students aged 19 to 21. The 44% of the students were male, and the 56% of them were female. The participants were informed about the purpose of the study, and they participated the study voluntarily by signing the consent form.

2.3. Data Collection Tools and Procedure

The Academic Dishonesty Tendency Scale developed by Eminoğlu and Nartgün (2009) was utilized to collect the data. The scale consists of 22 items measuring 4 latent dimensions. The first dimension named as "tendency towards cheating" includes 5 items, the second one, "dishonesty tendency at studies as homework" includes 7 items; the third dimension named as "dishonesty tendency at research and process of write up" has 4 items, and the last dimension, "dishonest tendency towards reference" consists of 6 items. The main reason of selecting this scale was that the issue of academic dishonesty had been investigated in detail in the scientific research courses lectured by the researchers. Another reason of preferring this scale within the context of the study was that the scale developers followed the main principles of the scale-development process neatly and provided the required reliability and validity evidence for the scale.

The scale development process began with literature review and analyzed undergraduate students' views towards academic dishonesty in terms of essays. At first draft of the scale, 40 items were written. The half of the items was negatively worded, while the other half of the items was positively worded. The items were presented to experts to get their ideas regarding the quality of the items, and based on the experts' suggestions, 15 items were discarded from the scale. The trial form of the scale was composed of 25 items. The respondents gave answers to the items on a 5 point-Likert scale (from 1 meaning "completely disagree" to 5 meaning "completely agree"). The trial form was administered to a sample including 300 participants. The psychometric qualities of the items and the whole scale were investigated based on different statistical techniques. The item-total correlations obtained for the items ranged from 0.27 to 0.68. The items were also analyzed based on the scores of the low and high group differences, and these differences were found significant for all items. The scale's construct validity was tested based on the exploratory and confirmatory factor analysis. The EFA was performed with the Principal Component Analysis and the Varimax method. The number of factors was determined based on the variance ratio and Kaiser criterion. The EFA revealed that the scale was composed of four dimensions, and the item loadings were between 0,558 and 0,743, with 53% explained variance ratio. (Eminoğlu & Nartgün, 2009).

The CFA was performed to be able to provide more evidence for the construct validity of the scale by the test developers. In the CFA, the t values of three items were found insignificant and discarded from the scale. The X2/sd ratio was found as 1.85, which provided evidence for a good model-data fit. All fit indexes were estimated as good levels, and the model-data fit was accepted as moderate and good level. Lastly, the reliability of the scale was investigated based on internal consistency. The test-retest and Cronbach Alpha coefficients were estimated for reliability of the scale, and both coefficients were found above 0.70. Based on these findings, the developers stated that the scale could be used to assess the academic dishonesty tendency of university students in a valid and reliable way (Eminoğlu & Nartgün, 2009).

In the current study, the scale was conducted on the study group during the two weeks of the fall semester of the 2018-2019 academic-year. The participation of the study group was voluntary. They were free to withdraw their consent for participation for any reason. In addition, they were informed about the goals of the study before the implementation of the scale.

2.4. Data Analysis

To gather evidence of validity, the CFA was performed to check whether the original factor structure of the scale was preserved in the present study or not. Firstly, the data were examined in terms of the assumptions of the CFA such as normality, multi-collinearity and singularity, linearity, missing and extreme values. The Maximum Likelihood Estimation method was preferred while carrying out the CFA because the normality assumption of the total score was met. Several fit statistics were also estimated to evaluate the model data fit. The Relative Chi Square Test, Root Mean Square Error of Approximation (RMSEA), Root Mean Square Residual (RMR), Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Relative Fit Index (CFI), Relative Fit Index (RFI), Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI) were considered while examining the model data fit. The related literature proposes various cut-off values for the result of the chi-square test, and the X2/df ratio. For example, Kline (2005) suggests that the values below 3 indicate perfect fit; the ones between 3 and 5 indicate a moderate fit. According to Brown's (2006) suggestions, the values $\leq .08$ are accepted good for the RMSEA, RMR and SRMR. The recommended thresholds indicating moderate values are mostly above 0.90 for the fit indices.

Secondly, the dimensionality of the data was analyzed based on the AISP method. At the first phase of the analysis, the exploratory Mokken scale analysis (Mokken, 1971) was used to examine the scalability and dimensionality of the scale. Furthermore, the scalability coefficients were estimated at this phase. The scalability coefficients were calculated at three levels: the item Hi, item-pair Hij, and scale level, H. Several lower bound values (c) for item level scalability coefficients (c=0.2 and c=0.3) have been proposed by researchers as lower bound values (Loevinger, 1948; Sijtsma & Molenaar, 2002). In the exploratory MSA, Hemker's procedure (Hemker et al., 1994) was adopted, and the AISP was used to select items to form scales. This procedure follows an iterative process. The homogeneous item clusters are composed at each step based on the scalability coefficients of the items, and the steps are repeated until no item satisfying the lower bound determined by the researchers remained. The H values start at 0 in the exploratory analysis and rise to 0.6 in 0.05 increments. In the current study, both the exploratory and confirmatory analyses were performed, and the AISP analyses were carried out on the R program by using the "mokken" package.

In addition to the AISP, the DETECT was also conducted to analyze dimensionality of the data. The exploratory and confirmatory DETECT analyses were carried out on the R program by using the "sirt" package. The confirmatory analysis was conducted based on the original structure explored by the scale developers. The index values (D, ASSI and Ratio) provided by the analyses for different item partitions were evaluated based on the criteria generally accepted for those index values. The D index value over 1 means that strong multidimensionality exists in the data. Index value between 0.40 and 1 indicates existence of medium level multidimensionality. Index value between 0.20 and 0.40 shows that weak dimensionality is observed in the data. Index values lower than 0.20 evidence that the data has an approximate simple structure. The ASSI (Approximate Simple Structure Index) and the Ratio index values could be accepted as the standardized forms of the DETECT index (Zhang, 2007). Similarly, the ASSI value over 0.25 and the ratio value over 0.36 indicate that the dataset shows significant deviation from the simple structure.

3. RESULTS

3.1. The Results Provided by the Confirmatory Factor Analysis

Firstly, the data were reviewed regarding the assumptions of the CFA. The Missing Completely at Random (MCAR) test was used to examine the missing values in the data. The results of the test yielded that the missing values occurred randomly. The 5 cases including missing values were discarded from the data set, and the CFA was performed on the 209 cases, which may be seen small for CFA. However, there are several studies proving that the sample size would be enough for the analysis. Some studies on the necessary sample size for the CFA noted considering the effects of the number of factors, the number of variables per factor and the size of communalities. The common conclusion of the related studies is that there cannot be a rule of thumb that can fit to every situation when deciding the sample size in the CFA. However, Monte Carlo simulation studies provided some guiding results on this issue. Mundrom et al. (2006) revealed that with a variables-to-factors ratio of at least 7, the minimum necessary sample size for excellent agreement is never greater than 180 and, in most cases, less than 150. Similarly, Wolf et al. (2013) revealed that if the number of variables per factor is equal to or higher than 6 necessary sample size does not exceed 200, even for the condition of low communalities. The scale utilized in the current study includes 4 factors having high variables-to-factors ratios. The numbers of the factor included by the 4 factors are 5, 7, 4 and 6, respectively. In addition, most scale items have loadings above 0.55, which indicates high communalities among items belonging to the same factors. Therefore, based on the findings of the related studies, the sample size of 209 can be accepted as enough for conducting CFA on the dataset.

The CFA was conducted to check whether the original four-dimensional structure of the scale was preserved in the current study or not. The results of the CFA revealed that the data obtained in the present study confirmed the original factorial structure of the scale. All fit indices indicated that the proposed model (four-dimensional model) yielded excellent or good model data fit [$\chi^2_{(203)}$ =428.98, p=.34; χ^2/df= 2.09; RMSEA=0.057 (0.049, 0.064; 90% CI); CFI=0.96; RFI=0.92; NFI=0.96; NNFI=0.96; GFI=0.90; AGFI=0.87; SRMR=0.058].

The standardized coefficients of the proposed model ranged from 0.40 to 0.82, above the lower bound value, 0.4 (Crocker & Algina, 1986). When the t-values of the items were analyzed, all of them were found significant, which evidence that all observed variables can be predicted by their latent variables. In addition to the item coefficients, the whole model was found significant in the assessment of academic dishonesty tendency of undergraduate students. Hence, the original factorial structure of the scale was preserved in the current study.

3.2. The Results Provided by the AISP

The exploratory MSA was preferred, and the scalability coefficients were calculated at the item, item pair and scale levels to investigate the suitability of the items to the Mokken scaling. The H_{ij} values were calculated for all item pairs, and it was revealed that all H_{ij} values were positive, and significantly higher than 0, which is the first requirement of the Mokken scaling. In the second step of the analysis, the item level scalability coefficients, H_i, were analyzed, and the Hi values estimated for the items were presented in Table 1. The item level scalability (H_i) coefficients given in Table 1 revealed that only three items (9, 12, and 15) had higher H_i coefficients than the lower bound value, c= 0.3. The low item scalability coefficients indicated that these items do not fit to a unidimensional structure. The scale level scalability coefficient (H) was found as 0.26, which supported that the scale is too weak to be scaled as a unidimensional scale. Upon estimating the scalability coefficients, the significances of these coefficients were analyzed, and all coefficients were found significant. Even though the items have low scalability values, the significance of the coefficients indicated that the MSA procedure may be applied.

Table 1. *The item level scalability coefficients - H_i value.*

Items	H_i coefficients	Standard error of H_i	Items	H_i coefficients	Standard Error of H_i
1	0.285	0.029	<u>12</u>	<u>0.313</u>	<u>0.029</u>
2	0.290	0.029	13	0.232	0.030
3	0.262	0.027	14	0.298	0.030
4	0.182	0.032	<u>15</u>	<u>0.307</u>	<u>0.031</u>
5	0.325	0.028	16	0.247	0.030
6	0.172	0.035	17	0.214	0.030
7	0.237	0.032	18	0.262	0.031
8	0.218	0.031	19	0.231	0.031
<u>9</u>	<u>0.309</u>	<u>0.029</u>	20	0.296	0.030
10	0.284	0.028	21	0.151	0.031
11	0.228	0.033	22	0.251	0.031
H value =	0.26				

The AISP procedure was started with the lowest value, c = 0.0. The AISP results obtained based on the c value of 0.0 revealed that all items were grouped into the same cluster as stated by Hemker et al. (1995), which was an expected finding. However, the c value of 0.0 should be accepted as a starting value, increasing gradually. It is suggested to try different lower bound values while scaling the items based on the AISP (Hemker et al., 1993; Meijer & Baneke, 2004). Depending on this suggestion, in the second step, the cut-off value for the AISP analysis was accepted as 0.2 and the obtained results were given in Table 2.

Table 2. *The results of the AISP.*

Items	Dimension Number	Items	Dimension Number
1	1	12	1
2	1	13	1
3	1	14	1
4	0	15	1
5	1	16	1
6	0	17	1
7	1	18	1
8	1	19	1
9	1	20	1
10	1	21	0
11	1	22	1

Table 2 indicated that the number of the dimensions for most items was defined as 1 by the second AISP analysis. This finding revealed that all items could compose a unidimensional scale, except for the three items (item 4, 6, and 21). These results evidenced that the scale could be accepted as unidimensional if the three items were excluded from the scale. However, the c value of 0.2 may lead to a weak scale because the lower scalability values will result in higher Guttman errors. Molenaar and Sijtsma (2000) proposed that the H values should be higher than 0.3 to get a reliable scale. In addition, the original factorial structure of the scale was multidimensional, and the CFA analysis of the data of the current study also confirmed the original four-dimensional structure. Therefore, the AISP was reiterated several times with higher cut-off values, c=2.25, 2.50, 2.75 and 3.00. The c values of 2.25, 2.50 and 2.75 provided similar results with each other, and the results were presented in Table 3.

Table 3. *The classifications of items according to AISP results.*

Items	Dimension Number			Items	Dimension Number		
	c values				*c* values		
	0.225	0.250	0.275		0.225	0.250	0.275
1	1	1	1	12	1	1	1
2	1	1	1	13	1	1	1
3	1	1	1	14	1	1	1
4	0	0	0	15	1	1	1
5	1	1	1	16	1	1	1
6	0	0	0	17	2	2	2
7	1	1	0	18	1	1	2
8	1	0	0	19	2	2	2
9	1	1	1	20	1	1	1
10	1	1	1	21	2	2	0
11	1	1	1	22	1	2	2

In Table 3, the numbers (0, 1, 2, and 3) indicated the number of possible dimensions of the scale. In addition, the numbers indicated the order of the dimensions, that is, the dimension number 1 meant that the items having this value belonged to the first dimension of the scale. Similarly, items having dimension numbers of 2 and 3 formed the second and the third dimension of the scale, respectively. The number 0, however, meant that these items had very low scalability coefficients, and the scalability criterion was not met for these items. It was found that for the *c* value of 0.225, 17 out of 22 items form a unidimensional scale, while 15 items constituted a unidimensional scale for the *c* value of 0.25. Lastly, 13 items out of 22 items formed a unidimensional scale for the c value of 0.275. The results also revealed that the number of the items that should be omitted from the scale increased as the *c* values got higher. In addition, the number of items included in the second scale increased based on the *c* values. These findings indicated that the scale has a multidimensional structure. The AISP was carried out again with different *c* values (0.300,0.325, 0.350 and 0.375), and the results were given in Table 4.

Table 4. *The classifications of items according to the second AISP results.*

Items	Dimension Number				Items	Dimension Number			
	c values					*c* values			
	0.300	0.325	0.350	0.375		0.300	0.325	0.350	0.375
1	1	1	1	1	12	1	1	1	2
2	1	1	1	1	13	0	0	2	2
3	1	1	1	1	14	1	1	2	2
4	0	0	0	0	15	1	1	2	2
5	1	1	1	1	16	0	2	3	3
6	3	4	0	0	17	2	3	4	4
7	0	0	0	0	18	2	3	4	4
8	0	0	0	0	19	2	3	4	4
9	1	1	3	3	20	1	1	2	2
10	1	2	3	3	21	0	0	0	0
11	3	4	0	0	22	2	3	4	0

Table 4 indicated that the dimensionality results obtained for the *c* values of 0.300, 0.325, 3.50, and 0.375 provided different results than the results obtained from the previous analyses carried out for the *c* values lower than 0.300. For example, the three dimensions were detected even for the *c* value of 0.300. The findings also revealed that the items grouped in the first scale were almost same for all *c* values. The items grouped in the first scale for the *c* values of 0.300 and 0.325 included item 1, 2, 3, 5, 9, 14, 15 and 20. In addition, the items 4, 7, 8, 11, 13 and 21 were detected as unscalable for more than one *c* value. The second, third and fourth dimensions included the items varied for each *c* value. These results confirmed that the scale has a multidimensional structure. However, the cluster partitions of the items are not consistent across the *c* values. Because of these inconsistences, the AISP was reiterated for the c values of 0.4, 0.425, and 0.450. The results suggested a seven-dimensional structure including fewer items, which is not applicable for the scale. Therefore, it was concluded that the results obtained from the analyses carried out for the c values of 0.350 and 0.375 were more similar to the scale's original factorial structure.

When the results obtained from the AISP were compared with the original factor solution achieved by the scale developers, it was seen that the item allocations were so different from the original scale at the all *c*-levels. The results obtained for the *c* value of 0.350 were accepted as the final result by taking into consideration Hemker et al.'s (1995) recommendations. For this c value, the four-factor solution was detected more balanced item distribution of scale's dimensions than the other c values. This item distribution pattern produced the most similar results with the original factor structure of the scale. In this solution, several items (item 4, 6, 7, 9, and 21) were not grouped under any factor. Based on these results, it was decided to discard these items from the scale. To summarize, the stepwise applications of the AISP indicated that the scale has a multidimensional structure, and the factor solution obtained for the c value of 0.350 can be accepted as the result of the AISP. However, it should be noted that this solution is not the same with the original factor solution proposed by the scale developers. It is the most similar one with the four-factor solution, but it proposed to discard 5 items from the scale, which resulting in the biggest difference from the original factor scale. When the items' distribution was analyzed, it was detected six items (I12, I15, I20, I9, I10 and I22) were allocated to the different factors from the original solutions. The other 10 items were estimated at the right factors as proposed by the original scale. This is the best solution created by the AISP; hence these results were accepted as the final solution for this technique.

3.3. The Results Provided by the DETECT Analysis

The exploratory DETECT analyses were carried out to analyze whether the dataset has simple structure or not. The index values estimated by the exploratory analysis for different item partitions were given in Table 5.

Table 5. *The results obtained from the exploratory DETECT analysis.*

The Number of Clusters	The D index	The ASSI	The Ratio
2	2.589	0.030	0.242
3	7.664	0.506	0.717
4	8.076	0.524	0.756
5	8.729	0.610	0.817
6	8.406	0.593	0.787
7	8.392	0.593	0.785

Table 5 indicated that the highest *D* index was estimated for the five-dimensional structure. The *D* index gives information regarding the structure of the data and the amount of multi-dimensionality observed in the data. A low index value means that inter-item covariances estimated conditioned on total scores are not high. This finding indicates that one dominant dimension explains inter-item relations and the dataset has a simple structure. A high index value shows that the dataset has a multidimensional structure. The *D* index value over 1 means strong multidimensionality exists in the data. According to the values given in Table 5, all of the *D* index values estimated for different item partitions were over 1. When the dataset was not unidimensional, obtaining a high *D* index value was expected since high conditional covariances among items belonging to the same item cluster. Therefore, the high DETECT, ASSI and Ratio index values given in the Table 5 revealed that conditional covariances among items were high. There were more than one dominant dimension explaining inter-item covariances, and the dataset showed significant differences from the unidimensional structure.

If the *D* index value evidences that the dataset has a multidimensional structure, it is necessary to specify the number of dimensions explaining the variance observed in the data and to explore how the items spread into different item clusters. The DETECT analysis estimated the highest *D* index for the five-dimensional structure. However, the original scale had a four-dimensional structure, and also the CFA results of the current study confirmed the original structure. Therefore, the confirmatory DETECT analysis was carried out based on the original structure defined by the scale developers. The index values provided by the analyses were given in Table 6.

Table 6. *The index values estimated by the exploratory and confirmatory DETECT.*

DETECT Analyses	Number of item cluster	D Index	ASSI	Ratio
Exploratory	5	8.729	0.610	0.817
Confirmatory	4	8.466	0.593	0.792

According to indices given in Table 6, the exploratory DETECT analysis indicated that the dataset obtained from applying the scale on the study group had five-dimensional structure. As stated before, the highest index values were estimated for the five-dimensional structure. The values calculated for the five-dimensional structure by the exploratory DETECT analysis were used as criterion to compare the results provided by the confirmatory DETECT analysis. The *D*, ASSI and Ratio index values estimated for the four-dimensional structure by the confirmatory analysis were high. The high values produced by the confirmatory analysis supported the results provided by the exploratory analysis. The results of both analyses indicated that the dataset has a multidimensional structure. When the index values were analyzed, it could be seen that the values obtained for the four-dimensional structure were very close to the values calculated for the five-dimensional structure. The cluster solution provided by the exploratory DETECT for the four-dimensional structure was given in Figure 1.

Figure 1. *The cluster solution provided by the exploratory DETECT.*

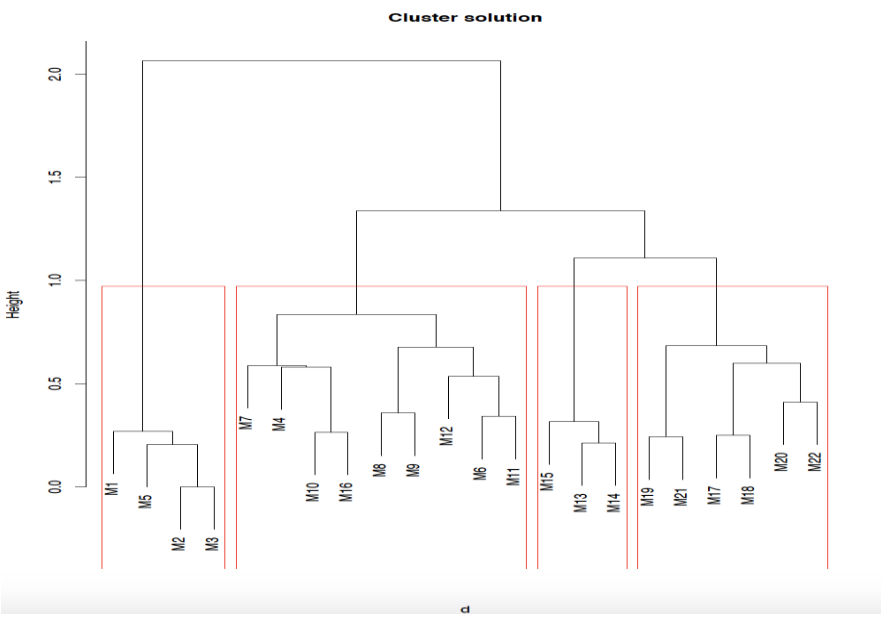


Figure 1 represents the cluster solution of the DETECT analysis. According to Figure 1, the first dimension included four items (1, 2, 3 and 5). The second dimension consisted of 9 items

(4, 6, 7, 8, 9, 10, 11, 12, 16), the third dimension included three items (13, 14, 15), and the fourth dimension consisted of six items (17, 18, 19, 20, 21, 22). To summarize, the results of both exploratory and confirmatory DETECT analyses indicated that the dataset is multidimensional, and items comprise homogenous item clusters. To make the results clearer and understand the proposed dimensionality structure for the scale, the obtained results from both techniques were given in Table 7.

Table 7. *The dimensionality structures proposed by the AISP and DETECT and the original scale.*

	Dimension1	Dimension2	Dimension3	Dimension4
AISP	1, 2, 3, 5, 12	9, 10, 16	13, 14, 15, 20	17, 18, 19, 22
DETECT	1, 2, 3, 5	4, 6, 7, 8, 9, 10, 11, 12, 16	13, 14, 15	17, 18, 19, 20, 21, 22
Original Scale	1, 2, 3, 4, 5	6, 7, 8, 9, 10, 11, 12	13, 14, 15, 16	17, 18, 19, 10, 21, 22

In Table 7, it is possible to see the items' allocation to the dimensions according to both techniques. Compared the results of the techniques with the original scale structure, it is clear that DETECT produced nearly the same factorial structure with the original scale. Only two items were placed to a different dimension, the other 20 items were However, as for AISP, the results were found so different from the original scale. Firstly, six of the 22 items were discarded from the scale based on the results of the AISP. The other dimensions suggested by the AISP were found similar with the other techniques, but the second dimension were found so different. Based on these results, it can be deduced that DETECT produced more suitable results with the original scale structures than the AISP.

4. DISCUSSION and CONCLUSION

The AISP analyses proposed several different factorial solutions. Firstly, the scalability coefficients were analyzed, and all coefficients were found low, but significant. In the MSA, the scalability coefficients have critical importance, and have been described as a method for evaluating a variety of measurement properties such as unidimensionality and local independence (Lind, 2017; Meijer et al., 2015). Despite of the recommended interpretations of scalability coefficients, there have been ongoing discussions regarding the usage of scalability coefficients in dimensionality analyses (Smits et al., 2012). That is why it may not be suitable to decide on the dimensionality of the scale based on the weak scalability coefficients. Therefore, the dimensionality of the data was examined by the AISP method. The related studies on scalability coefficients criticize that the accepted benchmarks (0.3) for the coefficients are so high that it is difficult to obtain these values for many scales, and items (Mokken & Lewis, 1982). For that reason, as stated by Hemker et al. (2015), the AISP analyses were reiterated for various *c* values to have more reliable evidence regarding the factorial structure of the scale.

Various *c* values ranging from 0.2 to 0.450 were utilized to reach the original factor solution of the data. When the c value of 0.2 was accepted as a cut-off value, it was found that the scale could be accepted as unidimensional except for 3 items. The *H* value was estimated as 0.26, which indicates a high Guttman error. Therefore, this solution was not acceptable for the scale. The AISP analyses were reiterated for the *c* values of 0.3, 0.325, 0.350, and 0.375. In addition, the results obtained from the AISP analyses carried out for the *c* values of 0.4, 0.425, and 0.450 were examined. However, it was concluded that the results of these analyses are too ambiguous to interpret. Furthermore, the results of these analyses suggested to discard several items from the scale, which might affect the content validity of the scale negatively. Although the results provided by the analyses are somehow inconsistent, it is still easy to infer from the results that the scale has a multidimensional structure.

The complex factor solutions in which items are mixed across the factors are generated by the AISP, when factors of scales are correlated with each other (Meijer & Baneke, 2004). In the current study, the AISP proposed several different and complex factorial structures for the scale with some unscalable items. In addition, the results of the AISP varied across the different c values. Because of the inconsistency among the results, it was concluded that the AISP may not be able to provide correct factor solutions in case that the scale has a multidimensional structure, and the correlations among these dimensions are medium or high levels (in this study, the inter-factors correlations ranged from 0.42 to 0.58).

In addition to the AISP, the dimensionality of the data was also examined based on the DETECT analysis. Similar with the AISP and the CFA results, the exploratory DETECT analysis supported the multidimensional structure of the scale. However, the highest index value was obtained for the five-dimensional structure by the exploratory DETECT analysis, while the CFA and AISP provided four and two-dimensional solutions, respectively. The exploratory DETECT analysis provided similar findings in terms of detecting the existence of the multidimensionality with the two methods, but the methods resulted with different factorial solutions. However, the exploratory DETECT analysis provided very similar cluster solution with the CFA. Only two items (4 and 16) were defined in different clusters by the two methods. While the DETECT revealed that item 4 belonged to the second dimension, the same item belonged to the first dimension in both the original-factorial structure and the structure defined in the current study. Similarly, the DETECT defined that item 16 belonged to the second dimension, while this item belonged to the third dimension in both the original and current study. The exploratory DETECT analysis provided results supporting the validity of four-dimensional structure explored by the CFA.

Similar with the AISP, both the exploratory and confirmatory DETECT analyses supported the existence of multidimensionality in the data. However, it is not possible to state that the AISP and DETECT analysis provided similar results regarding the factor numbers. The AISP defined four dimensions, while the DETECT analyses defined five factors underlying the scale items. In addition, the two methods provided very different item cluster solutions. The results of the analyses revealed that the DETECT provided more similar results with the CFA. The findings provided by the AISP were not in line with the factor solution proposed by the scale developers.

The results of the AISP analyses indicated that the scale is not suitable to be scaled based on the NIRT approach, which requires unidimensionality. It can be scaled based on the NIRT only if several items are excluded from the scale, but this situation may create new validity problems. Therefore, it is possible to state that the results obtained from the AISP did not support the original results of the scale. However, the AISP enabled to reveal multidimensionality observed in the data. The inconsistency between the factorial solution provided by the AISP and the original factorial structure might be caused by high correlations among dimensions of the scale. In the study conducted by Antion et al. (2018), the AISP correctly identified the dimensionality of the data, but in that study, the latent dimensions were uncorrelated. van Schuur (2003) mentioned the same drawback of the AISP. The researcher stated that in multidimensional scenario, only if the latent dimensions are uncorrelated, the AISP provides the accurate dimensionality. In addition, the results of the related studies (Antino et al., 2018; van Abswoude et al., 2004) confirmed van Schuur's (2003) claims. The findings of these studies revealed that correlations among latent dimensions result with relatively high Hij values for the items belonging to different dimensions, and the AISP erroneously tend to group all items in a single scale. The Hij values estimated in the current study ranged from 0.45 to 0.75, which indicates that there are medium and high correlations among the dimensions. As stated by Antino et al. (2018), the erroneous grouping effect often tends to occur wherever intermediate or high loading items are found together with moderately correlated latent dimensions. In addition, these situations may occur commonly in practice, therefore, the Mokken scale analysis may not be an adequate technique to explore the dimensionality of scales whose latent structure tend to be multidimensional. The results obtained from the AISP were consistent with the inferences of the study conducted by Antino et at. (2018). The scale utilized in the current study has a multidimensional structure, therefore the AISP could not be able to provide consistent results regarding the factorial structure of the scale. On the condition that the c value was accepted lower than the required level, the findings were found similar the findings reported van Abswoude et al. (2004) and Antino et al. (2018). They observed a tendency to lump items together in a single scale as in the findings of this study. Accordingly, it was concluded that it is necessary to utilize other dimensionality techniques together with the AISP when there is any suspicion regarding the existence of multidimensionality in the data.

Upon considering the related literature, it has been deduced that there is very limited number of studies investigating the usage of the AISP in the determination of the dimensionality. Wind (2017) stated that even though the AISP has been applied as a technique for evaluating the dimensionality and selecting items in affective domains, the usage of this procedure has not been fully explored especially in educational testing. The first study was conducted by Cavalini (1992), and the researcher compared the findings of factor analysis with the AISP. He used different lower bounds of scalability coefficients, and the results suggested that either three or four scales may be accepted. In the explanatory factor analysis, the four-factor solution was accepted as the best one. Hence, it may be accepted that the decisions about the number of dimensions should be made by considering reliability of the per scale score, the number of items in the per scale, and the interpretation of the meaning of the scales. Comparing the results of the EFA and AISP, the researcher deduced that the AISP can be used instead of the EFA in scale development process.

Another related study (Hemker et al., 1993) showed that results of the AISP may be affected by several factors such as the number of factors and correlations among factors. The number of items in separate factors may lead different solutions of the AISP. Considering these results, they proposed applying the AISP in the beginning of the scale development process. In addition, the researchers suggested that new studies should be done to compare the results obtained from empirical data sets and simulated data set together. To summarize, the results provided by the AISP in the current study, and the findings of the related research revealed that it is necessary to investigate the AISP method more to be able decide whether it is an effective dimensionality method or not.

The findings of the AISP did not provide the same factor solution proposed by the scale developers. However, both non-parametric methods (the DETECT and the AISP) revealed that the scale is multidimensional. Therefore, it is not appropriate to analyze the dataset based on the unidimensional IRT models. The results of the study indicated that both the DETECT and AISP succeeded to reveal the multidimensional structure of the scale. However, to determine the correct number of dimensions may not be the only goal in scale construction process. In this process, scale developers may want to create multidimensional scale of which factors are highly correlated. The AISP can provide strong evidence for the construct validity if the researchers select high cut-off values for the scalability coefficients.

The current study makes contributions from the methodological standpoint. In the first place, to the best of our knowledge, the present study is the first to compare the AISP and the DETECT with the CFA. On the other hand, our results obtained from the AISP are congruent with the findings reported by the related studies (Abswoude et al., 2004; Antino et al., 2018; Hemker et al., 1995). The researchers showed that the AISP may present misleading results when items and dimensions of scales have intermediate and high correlations among each other. In addition, we build on the existing work by showing the superiority of the parametric factorial techniques like the CFA compared to the non-parametric ones, such as the AISP and DETECT in the detection of the number of factors. Beyond the contribution made by the current study to the methodological literature, there are some practical implications of our findings for the researchers interested in social sciences. Our results revealed that the application of certain techniques under inadequate conditions may lead to erroneous results. Using only non-parametric techniques to examine dimensionality may cause researchers to make inaccurate decisions on the latent structures of the scale. To update the recommendations made by the related studies (Antino et al., 2018; van Abswoude et al., 2004; van der Eijk & Rose, 2015; van Schuur, 2003), social scientists are recommended to prefer the AISP only when the factorial structure is defined as unidimensional, or to develop a unidimensional scale. Another suggestion to the researchers regarding the AISP is to try out different lower bounds based on the item scalability coefficients. In a study by Meijer and Baneke (2004), conducting the AISP with a wide range of c values, it was found that if the item scalability coefficients are too low than the 0.3, high c values like 0.4 and higher may not produce meaningful results. For higher c values, the AISP generated so inconsistent results that the factorial solutions are almost impossible to interpret. In addition, researchers are advised to use the AISP method in dimensionality analysis only if the item scalability coefficients are higher than the lower bound values. As stated before, the AISP uses scalability coefficients based on the inter-item covariances, and if these coefficients are low, the AISP may generate inconsistent and unreliable results. Lastly, the usage of the DETECT analysis in combination with a parametric technique will provide more powerful and reliable results in examination of the dimensionality

Despite the theoretical and practical contributions of the current study, it is affected by several limitations that are discussed here together with the related future research. Firstly, the initial and whole item pool of the scale was not used in the dimensionality analyses process, since the scale was already developed, and the final version of it was available to use. This situation may have affected the results of the current study. Therefore, in the future studies, the researchers are recommended to use the DETECT and AISP techniques to analyze dimensionality by using the whole item pool, which may lead to different and more accurate results in terms of the factorial structure of the scale.

The second limitation of the study is that the correlations among dimensions were not manipulated, hence it might have altered the results as it was stated by the researchers (Antino et al., 2018; Hemker et al., 1993). In the future studies, correlations among dimensions may be controlled, and the effects of the correlations among factors on the dimensionality results can be observed. Thirdly, the item characteristics, such as item difficulty and discrimination indexes were not considered because the scale was already developed. Especially, item covariances may result with different factorial solutions in the AISP method, hence in the future studies, item covariances should be considered. Fourth, the data considered in the study was polytomous based on Likert response formats. However, dichotomous items are also used very frequently in educational settings. Therefore, researchers may examine dimensionality of the data obtained from dichotomous items. Lastly, the study group of the current study was relatively small, which may have affected the variances of the total scale scores, therefore, in the future research, the sample size can be modified to examine the factorial structure more neatly. For these reasons, in the future studies, these limitations should be addressed, and the researchers might apply several methods while deciding the number of factors. In that case, the results provided by the techniques may be more comparable, and both item characteristics and contents may be considered together in the process of the deciding the number of factors and items included in factors.

Declaration of Conflicting Interests and Ethics

The data of this study were gathered before February 2020. Hence, there is no ethical committee approval of the study. However, all ethical issues were taken into consideration during the data collection process by the authors. In addition, within the context of this study, the data were not used to make any decisions about the participants, only the theoretical comparisons were administered based on the results.

Authorship Contribution Statement

Ezgi MOR: Investigation, Resources, Analysis based on the automated item selection procedure, and Writing-original draft. **Seval KULA KARTAL**: Investigation, Analysis based on the DETECT, and Writing-original draft.

Orcid

Ezgi MOR https://orcid.org/0000-0003-0250-327X
Seval KULA KARTAL https://orcid.org/0000-0002-3018-6972

REFERENCES

Ackerman, T.A., Gierl, M.A., & Walker, C.M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(1), 37-53. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Antino, M., Alvarado, J.M., Asún, R.A., & Bliese, P. (2020). Rethinking the exploration of dichotomous data: Mokken scale analysis versus factorial analysis. *Sociological Methods Research, 49*(4), 839-867. https://doi.org/10.1177/0049124118769090

Cavalini, P.M. (1992). *It's an ill wind that brings no good. Studies on odour annoyance and the dispersion of odorant concentrations from industries* [Unpublished doctoral dissertation]. University of Groningen, The Netherlands.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Cengage Learning.

Finch, H. (2010). Item parameter estimation for the MIRT model bias and precision of confirmatory factor analysis based models. *Applied Psychological Measurement 34*(1), 10-26. https://doi.org/10.1177/0146621609336112

Finch, H. (2011). Multidimensional item response theory parameter estimation with non-simple structure items. *Applied Psychological Measurement, 35*(1), 67-82. https://doi.org/10.1177/0146621610367787

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(1), 255-282.

Guttman, L. (1950). *The basis for scalogram analysis*. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen (Eds.), Measurement and prediction (pp. 60-90). Princeton University Press.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139-164. https://doi.org/10.1177/014662168500900204

Hattie, J., Krakowski, K., Jane Rogers, H., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. *Applied Psychological Measurement, 20*(1), 1-14. https://doi.org/10.1177/014662169602000101

Hemker, B.T., & Sijtsma, K. (1993). A practical comparison between the weighted and the unweighted scalability coefficient of the Mokken model. *Kwantitatieve Methoden, 14*(44), 59-73.

Hemker, B.T., Sijtsma, K., & Molenaar, I.W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19*(4), 337-352. https://doi.org/10.1177/014662169501900404

Junker, B.W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *The Annals of Statistics, 21*(1), 1359-1378. https://doi.org/10.1214/aos/1176349262

Kane, M.T. (2006). Validation. *Educational Measurement, 4*(2), 17-64. https://doi.org/10.1111/j.1745-3992.1985.tb00874.x
Kim, H.R. (1994). *New techniques for the dimensionality assessment of standardized test data* [Doctoral dissertation, University of Illinois at Urbana Champaign]. ProQuest Dissertations and Theses Global.
Kuijpers, R.E., van der Ark, L.A. & Croon, M.A. (2013). Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. *Sociological Methodology, 43*(1), 42-69. https://doi.org/10.1177/0081175013481958
Lissitz, R.W. (Ed.). (2009). *The concept of validity: Revisions, new directions and applications*. Information Age Publishing.
Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects ofscale analysis and factor analysis. *Psychological Bulletin, 45*(1), 507-530. https://doi.org/10.1037/h0055827
Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison Wesley.
Meijer, R.R., Sijtsma, K., & Smid, N. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement, 14*(1), 283-298. https://doi.org/10.1177/014662169001400306
Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30*(10), 955 966. https://doi.org/10.1037/0003 066X.30.10.955
Mokken, R.J. (1971). *A theory and procedure of scale analysis*. De Gruyter.
Mokken, R.J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6*(1), 417-430. https://doi.org/10.1177/014662168200600404
Mokken, R.J., Lewis, C., & Sijtsma, K. (1986). Rejoinder to The Mokken scale: A critical discussion. *Applied Psychological Measurement, 10*(1), 279-285. https://doi.org/10.1177/014662168601000306
Molenaar, I.W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden, 3*(8), 145-164.
Molenaar, I.W. (1991). A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. *Kwantitatieve Methoden, 12*(37), 97-117.
Molenaar, I.W. (in press). *Nonparametric models for polytomous responses*. In W.J. van der Linden, & R.K. Hambleton (Eds.), Handbook of modern psychometrics (pp. 361-373). Springer.
Molenaar, I.W., Debets, P., Sijtsma, K., & Hemker, B. T. (1994). *User's manual for the computer program MSP* (Ver. 3.0). ProGAMMA.
Mroch, A.A., & Bolt, D.M. (2006). A simulation comparison of parametric and nonparametric dimensionality detection procedures. *Applied Measurement in Education, 19*(1), 67-91. https://doi.org/10.1207/s15324818ame1901_4
Nandakumar, R., & Ackerman, T. (2004). *Test modeling*. In D. Kaplan (Ed.), The SAGE handbook of quantitative methodology for the social sciences (pp. 93 107). SAGE Publications.
Reise, S.P., & Waller, N.G. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods, 8*(2), 164-184. https://doi.org/10.1037/1082-989X.8.2.164
Roussos, L.A., & Özbek, Ö.Y. (2006). Formulation of the DETECT population parameter and evaluation of detect estimator bias. *Journal of Educational Measurement, 43*(3), 215-243. https://doi.org/10.1111/j.1745-3984.2006.00014.x
Roussos, L.A., Stout, W.F., & Marden, J.I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35*(1), 1-30. https://doi.org/10.1111/j.1745-3984.1998.tb00525.x

Roznowski, M., Humphreys, L.G., & Davey, T. (1994). A simplex fitting approach to dimensionality assessment of binary data matrices. *Educational and Psychological Measurement, 54*(2), 263-283. https://doi.org/10.1177/0013164494054002002

Sick, J. (2010). *Assumptions and requirements of Rasch measurement*. SHIKEN: JALT Testing & Evaluation SIG Newsletter, 14(2), 23-29.

Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. SAGE Publications.

Sireci, S.G. (2009). *Packing and unpacking sources of validity evidence: History repeats itself again*. In R.W., Lissitz (Ed.), The concept of validity: Revisions, new directions and applications (pp. 19–37). IAP Information Age Publishing.

Slocum-Gori, S.L., & Zumbo, B.D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research, 102*(3), 443-461. https://doi.org/10.1007/s11205-010-9682-8

Stochl J, Jones P.B., & Croudace, T.J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology, 12*(1), 1-16. https://doi.org/10.1186/1471-2288-12-74

Stocking, M.L., Swanson, L., & Pearlman, M. (1991). *Automated item selection using item response theory*. Educational Testing Service.

Stocking, M.L., Swanson, L., & Pearlman, M. (1993). Application of an automated item selection method to real data. *Applied Psychological Measurement, 17*(2), 167-176. https://doi.org/10.1177/014662169301700206

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589-617. https://doi.org/10.1007/BF02294821

Stout, W. F. (1990). A new item response theory modelling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 49*(1), 293-325. https://doi.org/10.1007/BF02295289

Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika, 67*(4), 485-518. https://doi.org/10.1007/BF02295128

Stout, W., Froelich, A.G., & Gao, F. (2001). *Using resampling methods to produce an improved DIMTEST procedure*. In A. Boomsma, M.A.J. van Duijn, & T.A.B., Snijders (Eds.), Essays on item response theory (pp. 357-375). Springer.

Stout, W., Habing, B., Douglas, J., Kim, H.R., Roussos, L., & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20*(4), 331-354. https://doi.org/10.1177/014662169602000403

Stout, W., Nandakumar, R., & Habing, B. (1996). Analysis of latent dimensionality of dichotomously and polytomously scored test data. *Psychometrika, 23*(1), 37-65. https://doi.org/10.2333/bhmk.23.37

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*(1),159-203.

van Abswoude, A.A., van der Ark, L.A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. Applied Psychological Measurement, 28(1), 3-24. https://doi.org/10.1177/0146621603259277

van der Ark, L.A. (2007) Mokken scale analysis in R. *Journal of Statistical Software, 20*(11), 1-19. https://doi.org/10.18637/jss.v020.i11

van der Eijk, C. & Jonathan, R. (2015). Risky business: Factor analysis of survey data-assessing the probability of incorrect dimensionalisation. *PLoS One, 10*(3), 1-35. https://doi.org/10.1371/journal.pone.0118900

Wismeijer, A.A.J, Sijtsma, K., van Assen M.A.L.M & Vingerhoets, J.J.M. (2008). A comparative study of the dimensionality of the self-concealment scale using principal components analysis and mokken scale analysis. *Journal of Personality Assessment, 90*(4), 323-334. https://doi.org/10.1080/00223890802107875

Yu, C.H., Osborn Popp, S., DiGangi, S., & Jannasch Pennell, A. (2007). Assessing unidimensionality: A comparison of Rasch modeling, parallel analysis, and TETRAD. *Practical Assessment, Research, and Evaluation, 12*(1), 1-19. https://doi.org/10.7275/q7g0-vt50

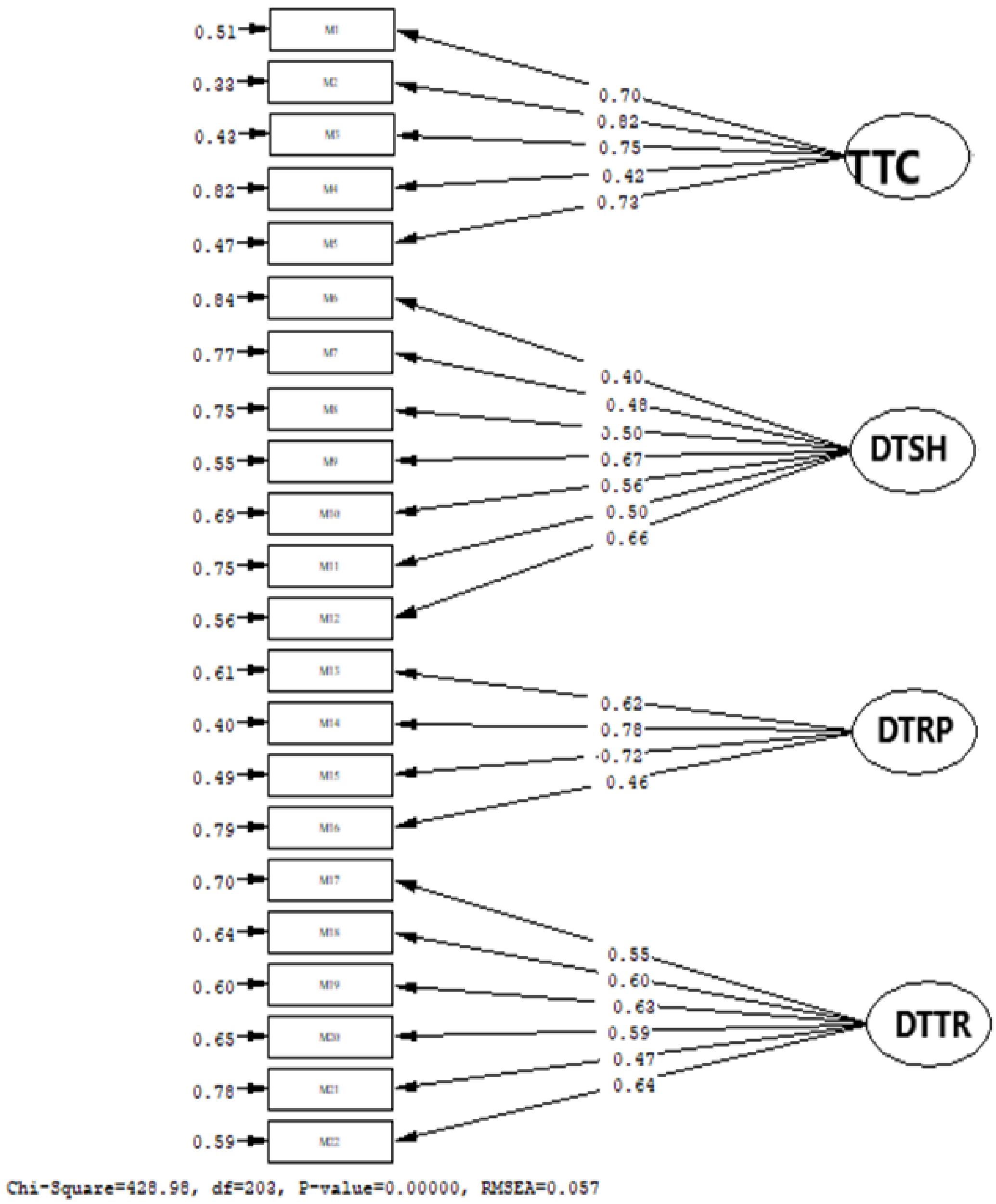
Yu, F., & Nandakumar, R. (2001). Poly Detect for quantifying the degree of multidimensionality of item response data. *Journal of Educational Measurement, 38*(2), 99-120. https://doi.org/10.1111/j.1745-3984.2001.tb01118.x

Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*(2), 213-249. https://doi.org/10.1007/ß

Zumbo, B.D. (2009). *Validity as contextualized and pragmatic explanation, and its implications for validation practice*. In R.W., Lissitz (Ed.), The Concept of Validity: Revisions, New Directions and Applications. IAP Information Age Publishing.

APPENDIX

The Path Diagram Provided by the CFA



TTP: Tendency towards cheating

DTSH: Dishonesty tendency at studies as homework

DTRP: Dishonesty tendency at research and process of write up

DTTR: Dishonesty tendency towards reference