Analyzing the Dimensionality of Academic Motivation Scale Based on the Item Response Theory Models*

Seval KULA KARTAL[1], Omer KUTLU[2]

A R T I C L E I N F O	A B S T R A C T

Purpose: This study aims to investigate the dimensionality of the Academic Motivation Scale items by depending on the graded response model, the generalized graded unfolding model, the bifactor model and the DIMTEST.

Research Methods: The Academic Motivation Scale was implemented on 1858 students who were studying at Ankara University. The fit of models was examined based on the general, person and item level model data fit statistics that were produced by the models.

Findings: It was found out that the bifactor model provided the most consistent results with the theoretical foundation underlying the items. The findings revealed that the generalized graded unfolding model and the bifactor model enabled better results than the graded response model concerning to the general model data fit. About item fit statistics, the models that provided the best fit were the bifactor model, the generalized graded unfolding model and the graded response model, respectively. The index values obtained based on the bifactor model also brought out the existence of a strong general dimension on which the scale items could be ordered. The results of DIMTEST analysis also supported that the scale items are multidimensional.

Implications for Research and Practice: Researchers are recommended to estimate item parameters both on the general dimension and subscales of the Academic Motivation Scale by utilizing the bifactor model to obtain more reliable and valid item parameter estimations. In future studies, researchers can compare the models about dimensionality and monotonicity assumptions based on scales developed to measure different affective traits.

Introduction

One of the fundamental aims of tests applied in the fields of education and psychology is to make deductions regarding the level of individuals' latent trait measured by the test (Lord, 1980; DeMars, 2010). To be able to make deductions regarding test-takers' trait levels, it is necessary to analyze the interaction between the level of the latent trait that the individual has and individual's responses to test items based on the mathematical models. The primary mathematical models that are used in the fields of education and psychology are developed based on the classical test theory (CTT) or the item response theory (IRT) under some assumptions such as dimensionality (Tate, 2002; Reckase, 2009; De Ayala, 2009). The statistical dimensionality indicates the minimum number of latent variables that is needed in order to summarize a matrix of item response data (Reckase, 1990). It means the necessary dimensionality to describe the interaction between individuals and items observed in the data matrix. The methods that are utilized to analyze the dimensionality of data and their assumptions determine the accuracy of the results regarding dimensionality. According to Tay and Drasgow (2012), when measurement models used to analyze the dimensionality of test data do not fit nature of the targeted latent trait, contradictory inferences are made regarding the dimensionality of data obtained from the application of an instrument tool. One of the measurement tools that conflicting deductions are made regarding its dimensionality is the Academic Motivation Scale (AMS), which is the focus of this study.

The Academic Motivation Scale (Vallerand, Pelletier, Blais, Brière, Senécal, & Vallières 1992) includes seven factors: three are related with intrinsic motivation, three are related with extrinsic motivation, and one measures amotivation. Each dimension has four items; therefore the scale includes 28 items. The Academic Motivation Scale builds upon the Self-Regulation Questionnaire, a well-known measure published for the first time by Ryan and Connell (1989) and the Self-Determination Theory. Since then, the measure has been adopted by researches exploring varied domains, including work motivation and academics. The Academic Motivation Scale is one example of the adaptation of the Self-regulation Questionnaire in the academic domain.

In the Self-Determination Theory, motivation is defined on the basis of three psychological needs as follows: competence, relatedness and autonomy. The theory argues different motivation forms of intrinsic or extrinsic motivation depending on these fundamental needs (Ryan & Deci, 2000a; 2000b). Motivation types differ from each other concerning the level of autonomy that they reflect. Therefore, different motivation types are regulated on a general continuum so that they can reflect various levels of autonomy (Deci & Ryan 2000; Viau 2009). For example, extrinsic motivation types locate on the left side, while intrinsic motivation types locate on the right side of the continuum. The level of autonomy that an individual has increases through the positive end of the continuum. Therefore, the theory suggests the existence of a one-dimensional continuum along which different motivation forms and items measuring these forms can lie from the negative end to the positive end of the continuum (Ryan & Deci, 2000a; Ryan, Rigby, & Przbylski, 2006).

The motivation structure defined in the Self-Determination Theory has been statistically examined by researchers, and the existence of the general dimension representing the autonomy continuum has been mostly supported by the studies based on the correlation analysis (Ryan & Connell, 1989; Vallerand et al., 1993; Fairchild, Horst, Finney, & Barron, 2005; Howard, Gagne, & Bureau, 2017). On the other hand, the studies in which dimensionality of the item response data was analyzed depending on the factor analytic methods revealed that seven-factor model better fit the data than one-factor model did (Vallerand et al., 1992; Fairchild, Horst, Finney, & Barron, 2005; Karagüven, 2012). The relationships pattern among subscales of the AMS providing results supporting one-dimensionality and factor analytic studies providing results supporting multi-dimensionality lead researchers to utilize more sophisticated statistical techniques to examine the motivation structure defined in the Self-Determination Theory.

More recent studies that analyze and examine the factorial structure of motivation were conducted based on the generalized graded unfolding model and bifactor modeling. The generalized graded unfolding model does not require the monotonicity assumption, which means that the probability of endorsing an item increases, or at least does not decrease, as the location of examinees increases on the latent trait dimension (Reckase, 2009). The bifactor model considers the possible multi-dimensionality that may be observed in the data, and allows for modeling both one and multidimensionality simultaneously. The studies in which the bifactor model was utilized evidenced the existence of a multidimensional motivation structure including both the general motivation factor and the group (or specific) factors reflecting different motivation types. In addition, it was found that factor loadings of items on the general motivation factor support the existence of the one-dimensional latent autonomy continuum (Gunnell & Gaudreau, 2015; Howard, Gagne, Morin, & Forest, 2016; Litalien, Morin, Gagne, Vallerand, Losier, & Ryan, 2017). The study utilizing the generalized graded unfolding model revealed that 18 items out of 28 items of the AMS fit the one-dimensional non-monotonic model (Miller, 2007).

The assumptions of the model may affect the results regarding the dimensionality of the data matrix. Therefore, it is very important to identify an appropriate measurement model allowing for considering different factors that may affect item responses of individuals, when analyzing dimensionality of a data matrix. For example, it was found that the monotonicity assumption might cause making incorrect inferences regarding the factor structure of measurement tools by negatively affecting dimensionality results (Tay & Drasgow, 2012). In addition, the related studies evidenced that item response models holding the monotonicity assumption are not always suitable to the nature of affective traits like academic motivation (Van Schuur & Kiers, 1994; Spector, Katwyk, Brannick, & Chen, 1997; Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001; Chernyshenko, 2003; Chernyshenko, Stark, Drasgow, & Roberts, 2007; Tay, Drasgow, Rounds, & Williams, 2009; Carter & Dalal, 2010; Tay & Drasgow, 2012; Cao, Drasgow, & Cho, 2015).

As explained before, the results of the studies utilizing the bifactor model indicated the existence of multidimensionality in the data obtained from answers provided by

students to the AMS. According to Tay and Drasgow (2012), it is necessary to analyze the data matrix concerning the monotonicity assumption by comparing model data fit of the monotonic and non-monotonic item response models, when results supporting the existence of multidimensionality are obtained for a data matrix that is expected to fit to a non-monotonic model. However, there are not many studies that examine the structure of the motivation data based on the non-monotonic item response model (Miller, 2007). In addition, there is not any study that analyzes fit of the AMS items to the monotonicity assumption by comparing model data fits of the monotonic and non-monotonic item response models.

Examining dimensionality of the data matrix obtained from the administration of the AMS based on different measurement models that hold different assumptions is significant to reach more valid and reliable results regarding dimensionality of the scale. Therefore, it is important to make decisions regarding dimensionality of the AMS items based on evidences obtained from sophisticated models developed under the item response theory. The reason of preferring item response theory over classical test theory in the current study is that the item response theory uses more information provided by the data since it allows for using whole response patterns of individuals rather than analyzing dimensionality based on only correlation or covariance matrix as factor analytic techniques do (Thissen & Wainer, 2001; Li, Jiao, & Lissitz, 2012). Along this line of research, the major purpose of the current study is to analyze dimensionality of the AMS items by utilizing the one and multi-dimensional item response models (graded response model (GRM), the generalized graded unfolding model (GGUM) and the bifactor model (BFM))

Method

Research Design

This research is a descriptive study in that this research provides information regarding fit of the AMS items to the monotonicity and dimensionality assumptions. This study is also a fundamental research aiming to examine the dimensionality of the data by comparing model data fits of different item response theory models.

Research Sample

The study group of this research consists of 1858 junior and senior students who were studying at the Faculties of Educational Sciences, Political Science, Communication, Engineering, Dentistry, Veterinary Medicine and Law of Ankara University during the Fall Term of the 2016-2017 Academic Year. 875 (47%) students were juniors while 983 (53%) of them were seniors. 726 (39%) out of 1858 students were male, 1132 (61%) of them were female students.

Research Instruments and Procedures

The data were obtained by conducting the Academic Motivation Scale in the study group. The scale was adapted from English to Turkish by Karagüven (2012). The exploratory and confirmatory factor analyses were carried out to examine the construct validity of the Turkish form of the scale. The confirmatory factor analysis

evidenced that model data fit statistics provided by the seven-factor structure of scale are acceptable, ($\chi^2 = 1017.74 \ (sd = 326, p < 0.05)$, $AGFI = 0.81, CFI = 0.94, SRMR = 0.065, RMSEA = 0.073$). The reliability and the construct validity of the AMS were also examined on the data obtained from the responses provided to the AMS by the participants of this study. Similar to the English form, results of the confirmatory factor analysis revealed that the seven factor-structure provided the best fit statistics among compared models (χ^2=3902.5 (sd=329, p<0.01), CFI=0.95, SRMR=0.07, RMSEA=0.07). The omega coefficient of the scale was calculated as 0.96. It is over the lower boundary that is accepted as 0.70-0.80 for the reliability (Reise & Revicki, 2015).

Data Analysis

Dimensionality of the data matrix obtained from the AMS was examined using the GRM, GGUM, BFM and DIMTEST analyses. The GRM and the BFM parameters were estimated on the R program using the "mirt" package (Chalmers, 2012), while the GGUM parameters were estimated on the GGUM2004 program (Roberts, Donoghue, & Laughlin, 2000). The GRM, GGUM and BFM were compared on the basis of the model fit statistics calculated at scale, person and item level.

Comparisons of item response theory models concerning the general model data fit statistics were carried out based on the Akaike (AIC), the Bayesian (BIC) and the adjusted Bayesian (A-BIC) information criteria (Li, Jiao, & Lissitz, 2012). To compare the GRM and the GGUM about their item fit, chi-square and degree of freedom ratios (χ^2/df) were calculated for item singlet, doublets and triplets (Carter, Guan, Maples, Williamson, & Miller 2015; Studts, 2008; Speer, Robie, & Christiansen, 2016). χ^2/df values were calculated on the MODFIT1.1 program (Stark, 2001). The S-χ^2 item fit statistics developed by Orlando and Thissen (2000) were calculated to compare the BFM with one-dimensional models concerning the item level model data fit. The "mirt" package on the R program was used to estimate the S-χ^2 statistic for the BFM. The GGUM2004 was used to calculate this statistic for the GGUM.

The l_z index value developed by Drasgow, Levine, and Williams (1985) was examined to compare models for their person level model data fit. The "mirt" package on the R program was used to calculate person fit statistics for the GRM and BFM. To calculate the l_z index value for the GGUM, a function developed by Tendeiro (2016) was adapted for the data of this study and run on the R program. Comparisons among models according to person level fit were made based on the mean of l_z values and the number of the individuals who were identified as unfit according to person fit statistics. Besides, the explained variance by the general and the group factors and the reliability coefficients were calculated based on the BFM estimations.

Results

The parameters for AMS items were estimated based on the GRM, GGUM and BFM. To enable the GGUM converge to the data, it was necessary to exclude two items of the AMS. Therefore, estimations were done for the remaining 26 scale items. The AIC, BIC, and A-BIC statistics calculated for the GRM, GGUM and BFM are given in Table 1.

Table 1

The General Model Data Fit Statistics

The models	AIC	BIC	A-BIC
GRM	151414.3	152420.3	151842.1
GGUM	128178.2	129324.4	128667.1
BFM	145636.9	146789.5	146125.7

Table 1 revealed that the AIC, BIC and A-BIC statistics calculated for the GGUM were lower than the ones estimated for the GRM. Thus, the general item fit statistics supported that the GGUM provided better model data fit than the GRM according to general fit statistics. It could be seen from Table 1 that the fit statistics calculated based on the BFM are lower than the ones calculated based on the GRM. Besides, it was found that the general model data fit of the BFM was significantly better than fit of the GRM according to -2 log likelihood values estimated for the models (χ^2=5829.4 (sd=26, p<0.05). The BFM achieved a 4% increase in the general model data fit. The GGUM had lower fit values than both the GRM and the BFM. Model comparisons based on the general fit statistics showed that the GGUM better fit the data than the model that took into consideration the multi-dimensionality The adjusted chi-square/degree of freedom (χ^2/df) ratios both for the GRM and the GGUM were calculated based on the response patterns given to the item singlets, doublets and triplets by the respondents to compare the item level model data fits of one-dimensional models. Mean, standard deviation and frequency distribution of χ^2/df ratios estimated for the models are given in Table 2.

Table 2

The Item Level Model Data Fit Statistics

Models	<1 f	1-2 f	2-3 f	3-4 f	4-5 f	5-7 f	>7 f	$\overline{X}$	σ
The GRM									
Singlets	0	0	0	0	0	0	26	1809.4	2961.8
Doublets	0	0	0	0	0	0	30	805.3	782.7
Triplets	0	0	0	0	0	0	14	81.9	38.9
The GGUM									
Singlets	24	1	0	0	0	0	1	2.4	11.4
Doublets	0	2	8	6	5	5	4	4.5	2.7
Triplets	0	1	6	3	2	2	0	3.5	1.2

In Table 2, the distribution of chi-square statistics was specified in columns that were called based on the size of the frequency value. For example, when Table 2 is examined according to the results provided by the GRM, the frequency values indicates that all of the chi-square statistics calculated for the 26 items of the scale are higher than 7. Therefore, it means that all of the adjusted χ^2/df ratios obtained for the item singlets, doublets and triplets based on the GRM are higher than the threshold value that was accepted as 3 for the item fit. High item fit statistics indicated that the GRM did not provide item level model data fit. When the chi-square statistics calculated for the GGUM were analyzed, it could be seen that the item fit statistics were lower than the ones calculated based on the GRM. Chi-square statistics estimated for the 25 items out of 26 items were lower than 2. The closeness of the χ^2/df ratios calculated for the item doublets and triplets to the threshold value indicated that the GGUM provided item level model data fit.

The S-χ^2 item fit statistics were calculated in order to compare the GGUM with the BFM concerning the item level model data fit. The mean of S-χ^2/df values for the GGUM was 1.87, while it was 1.07 for the BFM. The BFM provided better item level model data fit than the GGUM according to item fit statistics. It was accepted that the items whose S-χ^2/df value was over 3 do not fit to the model (Roberts, 2016). Therefore, it was found that all of the AMS items fit the BFM, while 24 items out of 26 items fit the GGUM. Examinations based on the adjusted chi-square/df ratios revealed that the GGUM provided better item level fit than the GRM. However, comparisons between the GGUM and the BFM showed that the BFM was the model that provides the best item level model data fit among the three models. The "l_z" person fit statistics were calculated based on the response pattern of individuals to compare the GRM, GGUM and BFM concerning their person level model data fit. The distribution and the mean of person fit statistics obtained for the models are presented in Table 3.

Table 3

The Person Level Model Data Fit Statistics

Models	$x \leq -4$ f	$-4 < x < -2$ f	$-2 \leq x < 0$ f	$0 \leq x \leq 2$ f	$2 < x < 4$ f	$x \geq 4$ f	$l_{z(Ort.)}$
GRM	36	141	527	1039	115	0	1.20
GGUM	981	552	270	53	2	0	5.76
BFM	601	411	536	304	6	0	3.41

GRM= Graded Response Model, GGUM= Generalized Graded Unfolding Model, BFM= Bifactor Model

According to the person fit statistics given in Table 3, 84.28% of the participants fit to the GRM, only 17.38% of them fit to the GGUM, 45.21% of them fit to the BFM. The GRM provided the best person level data fit among three models. According to the distribution of person fit statistics, the BFM is the second model that provided the best person level model data fit.

The explained common variance (ECV), the omega reliability coefficients and the PUC value were examined to compare the variance explanation power of the general dimension and the group factors (Reise, Moore, & Haviland, 2010; Qinn, 2015; Periard, 2016). Item parameters estimated based on the BFM were transformed to the factor analysis parameters (standardized factor loadings) to be able to calculate those values (Reckase & McKinley, 1991; Thissen & Wainer, 2001). The ECV, the PUC, and the omega coefficients calculated based on the general and group factor loadings of items are given in Table 4.

Table 4

Index Values Calculated Based on the Bifactor Model

	GD	S1	S2	S3	S4	S5	S6	S7
ECV	0.54	0.02	0.03	0.04	0.05	0.10	0.10	0.12
H	0.95	0.25	0.51	0.46	0.49	0.77	0.78	0.81
ω_S		0.90	0.80	0.83	0.81	0.83	0.85	0.93
ω_{HS}		0.09	0.12	0.20	0.33	0.58	0.69	0.46
PUC	0.89							
	0.96							
	0.85							

GD= General dimension, S1…S7= Subscale 1…Subscale 7, ECV= Explained common variance H= Structure reliability, ω_S= Omega coefficient for subscale, ω_{HS}= Hierarchical omega coefficient for subscale, ω_H = Hierarchical omega coefficient, PUC= Percent of uncontaminated correlation

The ECV values given in Table 4 showed the contribution of each dimension of the AMS to the explained variance. As can be seen from Table 4, the ECV value of the general dimension was 0.54. This value indicates that the general dimension itself explains 54% of the variance that is explained by the bifactor model. The generally accepted lower boundary for the ECV value is 0.60 (Reise, Scheines, Widaman, & Haviland, 2013; Periard, 2016). The ECV value over than 0.60 indicates the existence of a strong general dimension. The ECV value (0.54) calculated for the general dimension of the AMS was a little lower than this value. However, the PUC value should also be considered in order to evaluate the ECV value appropriately (Resie, 2012). The PUC value provides information regarding the number of correlations that were not affected by the existence of multi-dimensionality in the data matrix (Periard, 2016). According to Table 4, the PUC value is 0.89. Thus, it showed that 89% of correlation coefficients calculated among the AMS items were not affected by multi-dimensionality. Table 4 indicates that the hierarchical omega coefficient calculated for the general dimension is 0.85. The hierarchical omega coefficients calculated for subscales of the AMS range between 0.09 and 0.69. The hierarchical omega coefficients of the subscales were lower than the omega coefficients of the subscales. This finding

indicated that the reliability of subscales was increased by the effect of the general dimension.

The DIMTEST analysis was carried out to analyze dimensionality of the item response data matrix obtained from responses provided to the AMS based on a non-parametric method. It was found that all of the T statistics calculated when different items were used to form the assessment subtests (AT) were statistically significant. If AT items measure the same trait with the remaining scale items, it is expected to obtain low and statistically non-significant T-statistics. Having high and significant T-statistics means that the hypothesis that one dominant dimension can explain covariances among items included by AMS subscales was not confirmed.

Discussion, Conclusion and Recommendations

This study intended to examine the fit of AMS items to the monotonicity and dimensionality assumptions based on the GRM, the GGUM, the BFM and the DIMTEST analysis. Based on the general model data fit statistics, it was concluded that the GGUM provided the best model data fit, while the GRM provided the worst model data fit. The item level model data fit statistics of the GGUM showed that 23 items out of 26 items fit to the GGUM. However, according to item fit statistics, the GRM that is a monotonic one-dimensional model did not fit to any AMS item. The BFM was the model that provided the best item level model data fit among the three models. According to the person fit statistics, the GRM, the BFM and the GGUM provided the best person level model data fit, respectively

Based on the comparisons between the GRM and the GGUM according to the general and item fit statistics, it was found that the GGUM provided better item level and scale level model data fit than the GRM. Parallel with this finding, the result of Miller's study (2007) revealed that the GGUM fit to 18 items out of 28 AMS items. The researcher argued that the GGUM could be used as an alternative model to the confirmatory factor analysis that analyzes the data matrix based on a multi-dimensional approach.

The results of this study supported that the AMS items fit to tGGUM that does not assume monotone increasing item characteristics curves. Similar with this result, the studies conducted on instruments measuring various affective skills such as attitude and personality, revealed that the GGUM provided better model data fit than the monotonic models like the GRM (Roberts, Laughling, & Wedell, 1999; Chernyshenko, Stark, Chan, Drasgow, & Williams 2001; Chernyshenko, 2003; Meijer & Baneke, 2004; Chernyshenko, Stark, Drasgow, & Roberts, 2007; Miller, 2007; Cao, Drasgow, & Cho, 2015; Ling, Zhang, Locke, Li, & Li, 2016).

Dimensionality of the data matrix obtained from the responses given by participants to the AMS items was also examined based on the index values calculated depending on the item parameters estimated by the BFM. The hierarchical omega coefficient provided information regarding the total variance that can be attributed to the general dimension (Reise, Scheines, Widaman, & Haviland, 2013). Thus, 85% of the total score variance observed in the data was caused by the interpersonal differences

observed in the general motivation dimension of the AMS. The H-coefficients give information regarding the level of representation of a latent trait by its indicators and the level of re-attainability of this latent structure at different studies. The H-coefficients over 0.80 indicate that the specific latent trait is defined and represented well by its indicators (Reise, Scheines, Widaman, & Haviland, 2013). Therefore, H-coefficient estimated for the general dimension (0.95) indicated that the general motivation dimension was represented well by the AMS items, and the level of re-attainability of this structure at different studies is high.

The ECV, the PUC and the omega coefficients calculated to examine the power of general dimension and subscales of the AMS revealed that large percent of variance explained by the BFM was caused by the general dimension, and the reliability of subscales decreased when the effect of the general dimension was controlled. Although the AMS includes the seven subscales that measure different motivation types, items measuring these dimensions are mostly affected by the general dimension. Their degree of representing subscales to which they belong was low. Low ECV values of the subscales and high PUC value calculated for the scale supported the existence of a strong general dimension measured by the AMS. Parallel with this finding, the results of the study conducted by Litalien et al. (2017) revealed a general dimension measured by the AMS along which various motivation types (subscales of the AMS) indicating low or high levels of the autonomy can lie. Similarly, the results of Howard, Gagne, Morin and Forest's study (2016) on work motivation evidenced the existence of a well-defined general dimension reflecting one-dimensional autonomy continuum.

The index values revealed that the general motivation dimension caused large percent of variance that was explained by the BFM. This finding supported the existence of a strong general motivation dimension measured by the AMS. The highest omega reliability coefficients were calculated for the general dimension. When effects of the general dimension were controlled, very low reliability coefficients were obtained for the AMS subscales. Therefore, it could be stated that the BFM allowing the scale items to have loadings on both the general dimension and the subscales is the most convenient model to the multi-dimensional nature of the AMS items. Furthermore, results obtained based on the DIMTEST analysis revealed that the hypothesis of existence of one dimension explaining relationships among the AMS items was not confirmed. This finding indicated the existence of more than one latent trait that explained inter-item covariances of the AMS items. The results obtained based on the DIMTEST analysis supported the multi-dimensional nature of the AMS items.

Based on the model data fit statistics and the index values, it was concluded that the BFM provided the best fit to the items and the response patterns of participants among three models. Based on this result, researchers are recommended to estimate item parameters both on the general dimension and subscales of the AMS by utilizing the BFM to obtain more reliable and valid item parameter estimations. Similarly, instead of simply calculating the total scale or subscale scores, the researchers are suggested to estimate person parameters based on the BFM under item response

theory or factor analytic approach to more appropriately estimate motivation levels of the respondents.

This study examined the model data fits of the GGUM, the GRM and the BFM based on the responses given to the motivation scale items by the respondents. These models can be compared concerning dimensionality and monotonicity assumptions based on scales developed to measure different affective traits, such as attitude, personality. The BFM used in this study is a monotonic multi-dimensional model. In future studies, an item response theory model (multidimensional generalized graded unfolding model) that considers both the monotonicity and multi-dimensionality assumptions can be included and compared with the GRM, the GGUM and the BFM concerning model data fit. The current study was carried out on the data obtained from students studying at a university. The study group included a large sample of students from different faculties; however, it is still possible that students from different universities may follow different cognitive or psychological processes while answering items of the AMS. Therefore, the models can be compared within different samples of students.

References

Berkeljon, A. (2012). *Multidimensional item response theory in clinical measurement: a bifactor graded- response model analysis of the outcome questionnaire-45.2.* (Doctoral dissertation). Retrieved from Brigham Young University Scholars Archive.

Broadfoot, A. A. (2008). *Comparing the dominance approach to the ideal-point approach in the measurement and predictability of personality.* (Doctoral dissertation). Retrieved from OhioLINK Databases.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: The Guilford Press.

Cao, M., Drasgow, F., & Cho, S. (2015). Developing ideal intermediate personality items for the ideal point model. *Organizational Research Methods, 18*(2), 252-275.

Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI work satisfaction scale. *Personality and Individual* Differences, 49, 743-748.

Carter, N. T., Guan, L., Maples, J. L., Williamson, R.L., & Miller, J.D. (2015). The downsides of extreme conscientiousness for psychological wellbeing: The role of obsessive compulsive tendencies. *Journal of* Personality, 84(4), 510 522.

Chalmers, R. P. (2012). A multidimensional item response theory package for the R environment. *Journal of Statistical* Software, 48(6), 1-29.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189–225.

Chernyshenko, O. S. (2003). Applications of ideal point approaches to scale construction and scoring in personality measurement: The development of a six-faceted measure of conscientiousness. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 63*(11-B), 5556.

Chernyshenko, O. S., Stark, S. E., Drasgow, F., & Roberts, J. S. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*(1), 88-106.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*(4), 523-562.

De Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: The Guilford Press.

DeMars, C. (2010). *Item response theory.* New York: Oxford University Press, Inc.

Drasgow, F., Levine, M. F., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response model and standardized indices. *British Journal of Mathematical and Statistical Psychology, 38*, 67-86.

Embretson, S.E., & Reise, S.P.(2000). *Item response theory for psychologists.* New Jersey: Lawrence Erlbaum Associate, Inc.

Fairchild, A. J., Horst, S. J., Finney, S. J., & Barron, K. E. (2005). Evaluating existing and new validity evidence for the Academic Motivation Scale. *Contemporary Educational Psychology, 30*, 331-358.

Gibbons, R. D., & Hedeker, D. R. (1992). Full information item bifactor analysis. *Psychometrika, 57*(3), 423-436.

Gunnell, K. E., & Gaudreau, P. (2015). Testing a bifactor model to disentangle general and specific factors of motivation in self-determination theory. *Personality and Individual Differences, 81*, 35-40.

Howard, J. L., Gagne, M., & Bureau, J. S. (2017). Testing a continuum structure of self-determined motivation: A meta-analysis. *Psychological Bulletin, 143*(12), 1346-137.

Immekus, J., & Imbrie, P. K. (2008). Dimensionality assessment using the full information item bifactor analysis for graded response data an illustration with the State Metacognitive Inventory. *Educational and Psychological Measurement, 68*(4), 695-709.

Karagüven, H. (2012). Akademik Motivasyon Ölçeğinin Türkçeye adaptasyonu. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(4), 2599-2620.

Li, Y., Jiao, H., & Lissitz, R. W. (2012). Applying multidimensional item response theory models in validating test dimensionality: An example of K–12 large scale science assessment. *Journal of Applied Testing Technology, 13*(2), 2-27.

Ling, Y., Zhang, M., Locke, K. D., Li, G., & Li, Z. (2016) Examining the process of responding to circumplex scales of interpersonal values items: Should ideal point scoring methods be considered? *Journal of Personality Assessment, 98*(3), 310-318.

Litalien, D., Morin, A. J. S., Gagne, M., Vallerand, R. J., Losier, G. F., & Ryan, R. M. (2017). Evidence of a continuum structure of academic self-determination: A two-study test using a bifactor-ESEM representation of academic motivation. *Contemporary Educational Psychology, 51*, 67-82.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* NJ: Lawrence Erlbaum.

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods, 9,* 354- 368.

Miller, B. J. (2007). *Unfolding analysis of the Academic Motivation Scale: A different approach to evaluating scale validity and Self-Determination Theory* (Doctoral dissertation). Retrieved from ProQuest Digital Dissertations.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational and Behavioral Statistics, 18*(1), 41-68.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24,* 50-64.

Periard, D. (2016). A bifactor model of burnout? *An item response theory analysis of the Maslach burnout inventory* (Doctoral dissertation). Retrieved from OhioLINK Databases

Quinn, H. O. (2014). *Bifactor models, explained common variance (ECV) and the usefulness of scores from unidimensional item response theory analyses.* (Master's thesis). Retrieved from University of Carolina Digital Repository.

Reckase M. D, & McKinley, R. L. (1991) The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement 15,* 361-373.

Reckase, M. D. (1990). *Unidimensional data from multidimensional tests and multidimensional data from unidimensional tests.* Paper presented at the Annual Meeting of the American Educational Research Association.

Reckase, M. D. (2009). *Multidimensional item response theory.* New York: Springer.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667-696.

Reise, S. P., & Revicki, D. A. (2015). *Handbook of item response theory modeling.* New York: Taylor & Francis Group.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2013). Applying unidimensional item response theory models to psychological data. K. F. Geisinger (Ed.). *APA handbook of testing and assessment in psychology.* Washington DC: APA.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26.

Roberts, J. S. (1995*). Item response theory approaches to attitude measurement* (Doctoral dissertation). University of South Caroline, USA.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1999). *Estimating parameters in the generalized graded unfolding model: Sensitivity to the prior distribution assumption and the number of quadrature points used.* Paper presented at the Annual Meeting of the National Council on Measurement in Education

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3–32.

Roberts, J. S., Laughlin, J. E., & Wedell, D. H. (1999). Validity issues in the Likert and Thurstone approaches to Attitude Measurement. *Educational And Psychological Measurement, 59*(2), 211-233.

Ryan, R. M., & Connell, J. M. (1989). Perceived locus of causality and ınternalization: examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*(5), 749-761.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54-67.

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motiv Emot, 30*, 347-363.

Sgammato, A. N. (2009). *An application of unfolding and cumulative item response theory models for noncognitive scaling: Examining the assumptions and applicability of the generalized graded unfolding model* (Doctoral dissertation). Retrieved from University of Carolina Digital Repository.

Spector, P. E., Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristic can produce artifactual factors. *Journal of Management, 23*(5), 659-677.

Speer, A. B., Robie, C., & Christiansen, N. D. (2016). Effects of item type and estimation method on the accuracy of estimated personality trait scores:

Polytomous item response theory models versus summated scoring. *Personality and Individual Differences, 102,* 41–45.

Stark, S. (2001). *MODFIT: A computer program for model-data fit.* Urbana Champaign, IL: University of Illinois at Urbana-Champaign.

Studts, C. R. (2008). *Improving screening for externalizing behavior problems in very young children: Applications of item response theory to evaluate instruments in pediatric primary care* (Doctoral dissertation). Retrieved from ThinkIR: The University of Louisville's Institutional Repository

Tate, R. (2002). Test dimensionality. G. Tindal, & T. M. Haladyna (Ed.). *Validity, technical adequacy, and implementation* (p. 181-213). New Jersey: Lawrence Erlbaum Associates, Publishers.

Tay, L., & Drasgow, F. (2012). Theoretical, statistical, and substantive ıssues in the assessment of construct dimensionality: Accounting for the item response process. *Organizational Research Methods, 15*(3), 1-22.

Tay, L., Drasgow, F., Rounds, J., & Williams, B. (2009). Fitting measurement models to vocational ınterest data: Are dominance models ıdeal? *Journal of Applied Psychology, 94*, 1287-1304.

Tendeiro, J. N. (2016). The $l_{z(p)}$ person-fit statistic in an unfolding model context. *Applied Psychological Measurement, 41*(1), 44-59.

Thissen, D., & Wainer, H. (2001). *Test scoring.* NJ: Lawrence Erlbaum Associates, Inc.

Vallerand, R. J., Pelletier, L. G., Blais, M. R, Brière, N. M., Senécal, C., & Vallières, E. F. (1993). On the assessment of intrinsic, extrinsic and amotivation in education: evidence on the concurrent and construct validity of the academic motivation scale. *Educational and Psychological Measurement, 53*, 159-172.

Vallerand, R. J., Pelletier, L. G., Blais, M. R, Brière, N. M., Senécal, C., & Vallières, E. F. (1992). The academic motivation scale: a measure of ıntrinsic, extrinsic and amotivation in education. *Educational and Psychological Measurement, 52*, 1003-1017.

Van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead. *Applied Psychological Measurement, 18*(2), 97-110.

Akademik Güdülenme Ölçeği'nin Boyutluluk Açısından Madde Tepki Kuramı Modellerine Dayalı Olarak İncelenmesi

Atıf:

Özet

Problem Durumu: Eğitimde ve psikolojide kullanılan en temel matematiksel modeller bazı varsayımlara dayalı olarak geliştirilmiştir. Bu varsayımlarından biri ölçme araçlarından elde edilen madde yanıt matrisinin boyutluluğuna ilişkindir. Farklı modellerle ve test koşullarıyla boyutluluğun incelendiği yöntemler sonuçların doğruluğu üzerinde belirleyici olmaktadır. Ölçme modelleri örtük özelliğin doğasına uygun olmadığında, ölçme aracının boyutluluğuna ilişkin çelişkili çıkarımlar yapılmaktadır. Boyutluluğu açısından çelişkili bulguların elde edildiği ölçme araçlarından biri bu araştırmanın odak noktası olan Akademik Güdülenme Ölçeği (AGÖ)'dir.

Modellerin dayandığı varsayımlar boyutluluk yöntemleriyle elde edilen bulguları etkileyebildiğinden, boyutluluğuna ilişkin doğru bulgulara ulaşılması için Akademik Güdülenme Ölçeği maddelerine verilen yanıtları en iyi betimleyen ölçme modelinin belirlenmesi ve boyutluluğun bu modelle incelenmesi gerekmektedir. İlgili çalışmalar, monoton olmayan madde tepki kuramı modelinin duyuşsal özelliklerin doğasına daha uygun bir ölçme modeli olduğunu ortaya koymuştur. Ayrıca, monotonluk varsayımının boyutluluğa ilişkin bulguları etkileyerek ölçme araçlarının faktör yapılarına ilişkin hatalı çıkarımlara neden olduğu bulunmuştur. Bu nedenle, monoton olmayan madde tepki kuramı modeline uyduğu düşünülen bir veride çok boyutluluğun varlığına ilişkin kanıtlar elde edildiğinde, verinin monotonluk varsayımı açısından monoton ve monoton olmayan modellerin model veri uyumlarının karşılaştırılması yoluyla incelenmesi gerekmektedir. Ancak, alanyazında güdülenmenin monoton olmayan modele dayalı olarak incelendiği sınırlı sayıda çalışma bulunmaktadır. Ölçek maddelerinin monotonluk varsayımına uygunluğu açısından, monoton ve monoton olmayan MTK modellerinin karşılaştırıldığı bir çalışmaya ise rastlanmamıştır. Ayrıca, çalışmalarda AGÖ maddelerinden elde edilen yanıt matrisinin boyutluluğunun faktör analizi çerçevesinde incelendiği, boyutluluk incelemelerinin MTK çerçevesinde parametrik ve parametrik olmayan yöntemlerle yapılmadığı görülmüştür. Ölçek maddelerinin boyutluluğuna ilişkin kararların, MTK çerçevesinde geliştirilmiş parametrik ve parametrik olmayan yöntemlerle elde edilen daha kapsamlı kanıtlara dayalı olarak verilmesi önemli görülmektedir. Alanyazındaki bu eksiklikler doğrultusunda, AGÖ maddelerinin boyutluluk açısından MTK çerçevesinde geliştirilmiş tek ve çok boyutlu parametrik ve parametrik olmayan modellere dayalı olarak incelenmesi gerekli görülmektedir.

Araştırmanın Amacı: AGÖ maddelerinin boyutluluk açısından MTK çerçevesinde geliştirilmiş Aşamalı Tepki Modeli (ATM), Genelleştirilmiş Aşamalı Monoton

Olmayan Model (GAMOM), İki Faktör Modeli (İFM) ve DIMTEST analizine dayalı olarak incelenmesi araştırmanın amacını oluşturmaktadır.

Araştırmanın Yöntemi: Bu çalışma, AGÖ maddelerinin boyutluluk ve monotonluk varsayımlarına uygunluğu açısından bilgi sağlaması yönüyle betimsel, boyutluluğun farklı madde tepki kuramı modellerinin model veri uyumuna dayalı olarak karşılaştırılmasına dayalı olarak incelenmesi yönüyle de temel bir araştırma niteliğindedir. Araştırmanın çalışma grubunu 2016-2017 eğitim-öğretim yılı güz döneminde Ankara Üniversitesi'nin Eğitim Bilimleri, Siyasal Bilgiler, İletişim, Mühendislik, Diş Hekimliği, Veteriner ve Hukuk Fakültelerinde öğrenimini sürdüren üçüncü ve dördüncü sınıf öğrencileri oluşturmaktadır. Çalışma grubunda toplam 1858 öğrenci bulunmaktadır.

Araştırmanın verileri Akademik Güdülenme Ölçeği kullanılarak toplanmıştır. Fransızca olarak geliştirilen ölçeğin İngilizceye uyarlanması Vallerand ve diğerleri (1992) tarafından yapılmıştır. İngilizce formun yapı geçerliğinin incelenmesi amacıyla yapılan doğrulayıcı faktör analizi sonucunda yedi faktörlü yapı için kabul edilebilir uyum değerleri elde edilmiştir. Ölçeğin alt boyutları için elde edilen Cronbach Alfa güvenirlik katsayıları 0.83 ile 0.86 arasında değişmektedir. Akademik Güdülenme Ölçeği'nin İngilizce'den Türkçe'ye uyarlanması Karagüven (2012) tarafından yapılmıştır. Türkçe formun yapı geçerliğinin incelenmesi amacıyla açımlayıcı ve doğrulayıcı faktör analizi yapılmıştır. Doğrulayıcı faktör analizi sonucunda model veri uyumu değerlerinin yüksek olduğu ve özgün ölçeğin yedi faktörlü yapısının doğrulandığı bulunmuştur. Çalışmada ölçeğin boyutları; İçsel Motivasyon (Bilme, Başarma, Uyarım), Dışsal Motivasyon (Dışsal Düzenleme, İçe Yansıyan, Belirlenmiş) ve Motivasyonsuzluk olarak adlandırılmıştır.

Akademik Güdülenme Ölçeği'nden elde edilen madde yanıt matrisinin boyutluluğunun incelenmesi amacıyla Aşamalı Tepki Modeli (ATM), Genelleştirilmiş Aşamalı Monoton Olmayan Model (GAMOM) ve İki Faktör Modeli (İFM)'ne dayalı olarak madde ve birey parametre kestirimleri yapılmıştır. Ayrıca DIMTEST analizine dayalı olarak da boyutluluk incelemesi yapılmıştır. ATM'ye dayalı kestirimler R programında "mirt" paketi (Chalmers, 2012), GAMOM'a dayalı kestirimler GGUM2004 programı (Roberts, Donoghue ve Laughlin, 2000) kullanılarak yapılmıştır. İFM'ye dayalı parametre kestirimleri R programında "mirt" paketi (Chalmers, 2012) kullanılarak elde edilmiştir. Modeller arası karşılaştırmalar ölçek, madde ve birey düzeyinde genel, madde ve birey model veri uyumu istatistikleri hesaplanarak yapılmıştır. Ayrıca, iki faktör modeline dayalı olarak genel ve grup faktör tarafından açıklanan varyans oranları ve güvenirlik değerleri incelenmiştir.

Araştırmanın Bulguları: Genel model veri uyumu istatistiklerine dayalı olarak yapılan modeller arası karşılaştırmalar İFM'nin genel model veri uyumunun ATM'ye göre anlamlı biçimde daha iyi olduğu, ($\chi^2_{sd=26}$=5829.4 p<0.05) ve model veri uyumunda %4'lük anlamlı bir iyileşme sağladığı bulunmuştur. Ancak, GAMOM'un hem ATM hem de İFM'ye göre daha düşük genel model veri uyumu istatistiklerine sahip olduğu bulunmuştur.

ATM'ye dayalı olarak maddeler için hesaplanan tüm uyum istatistiklerinin model veri uyumu için kabul edilen sınır değerin çok üzerinde olduğu görülmüştür. Uyum istatistiklerinin yüksek olması, ATM'nin madde düzeyinde model veri uyumu sağlamadığını göstermiştir. AGÖ maddelerinin tamamının İFM'ye uyum sağladığı, iki maddenin ise GAMOM'a uyum sağlamadığı bulunmuştur. Madde uyum istatistikleri, ATM'ye göre GAMOM'un madde düzeyinde sağladığı model veri uyumunun daha iyi olduğunu göstermiştir. İFM ve GAMOM arasında yapılan karşılaştırmalara dayalı olarak ise İFM'nin madde düzeyinde en iyi model veri uyumu sağlayan model olduğu bulunmuştur. Birey uyum istatistiklerine göre, üç model içerisinde birey düzeyinde en iyi model uyumunu ATM sağlamıştır. Uyum istatistiklerinin dağılımına göre İFM, ATM'den sonra birey düzeyinde en iyi uyumu sağlayan ikinci model olmuştur.

Genel boyut ve alt boyutların madde-yanıt matrisini açıklamadaki gücünü karşılaştırabilmek amacıyla hesaplanan indeks değerlerine dayalı olarak ölçek maddeleri ile ölçülen güçlü bir genel boyutun var olduğu bulunmuştur. DIMTEST analizi sonucunda yüksek T-istatistikleri ve anlamlılık değerleri elde edilmiştir. Yüksek değerler tüm alt boyutlarda yüksek koşullu kovaryansların elde edildiğini göstermiştir. Ölçeğin alt boyutlarında yer alan maddeler arasındaki ilişkilerin, baskın bir boyut tarafından açıklanabileceği hipotezinin doğrulanmadığı bulunmuştur.

Araştırmanın Sonuçları ve Önerileri: ATM, GAMOM ve İFM için model veri uyumu istatistiklerine ve indeks değerlerine dayalı olarak yapılan karşılaştırmalar birlikte düşünüldüğünde, ATM'ye göre monotonluk varsayımında bulunmayan tek boyutlu model ölçek maddelerine daha uyumludur. Ancak, üç model içerisinde İFM'nin ölçek maddelerine ve bireyler tarafından maddelere verilen yanıt örüntülerine en uyumlu model olduğu sonucuna ulaşılmıştır. Bu bulguya dayalı olarak araştırmacılara, bireylerin güdülenme düzeylerine ilişkin daha geçerli ve güvenilir sonuçlar elde etmek için hem genel boyutta hem de alt boyutlarda birey parametre kestirimlerini İFM'ye dayalı olarak yapmaları önerilmektedir. Benzer biçimde, Akademik Güdülenme Ölçeği'ni kullanmak isteyen uygulayıcıların da ölçekten toplam puan ya da alt boyut puanları hesaplamak yerine, bireylerin güdülenme düzeylerini belirlemek için madde tepki kuramı ya da faktör analizi çerçevesinde iki faktör modellemesine dayalı olarak kestirim yapmaları önerilmektedir.

Çalışma kapsamında, monoton ve monoton olmayan tek ve çok boyutlu modellerin uyumu güdülenme ölçeğine verilen yanıtlara dayalı olarak incelenmiştir. Tutum, ilgi gibi farklı duyuşsal özellikleri ölçmek amacıyla geliştirilmiş ölçekler üzerinde ATM, GAMOM ve İFM kullanılarak monotonluk ve boyutluluk varsayımları açısından karşılaştırmalar yapılabilir. Bu araştırmada kullanılan İFM de monotonluk varsayımına dayanan bir modeldir. Bu nedenle, hem çok boyutluluğu hem de monotonluk varsayımını göz önünde bulunduran çok boyutlu monoton olmayan MTK modeli de dahil edilerek modeller model veri uyumları açısından karşılaştırılabilir.

Anahtar Sözcükler: Çok boyutlu madde tepki kuramı, iki faktör modeli, genelleştirilmiş aşamalı monoton olmayan model, boyutluluk ve monotonluk varsayımları.