ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2024; 15(1); 18-34

Measurement Invariance of Achievement Motives Model: PISA 2018 Turkey Sample

Münevver BAŞMAN*

Abstract

This study aims to find out whether the achievement motives model, constructed by attitudes towards competition, motivation to master tasks, and fear of failure scales, has measurement invariance in the PISA 2018 student questionnaire concerning gender and school type in Turkey sample, containing 6442 students. According to the results, the model's fit levels with the data were within acceptable levels across gender groups and school-type groups. Then, the measurement invariance across gender and school type was tested by multigroup confirmatory factor analysis including a sequence of tests of four nested hierarchical models which are configural, metric, scalar, and strict invariance. The fit indices of models and the differences of indice values between models were examined to decide whether measurement invariance is established. It is found that the full measurement invariance holds according to gender and school type since the values of the indices for each invariance step are acceptable. It means that it will be appropriate and meaningful to compare the students based on the scores obtained from the achievement motives model.

Keywords: achievement motives, gender, measurement invariance, PISA, school type

Introduction

International assessments allow countries to observe their successes and shortcomings as well as their situation compared with other countries. One of these international assessment studies conducted in this direction is the Programme for International Student Assessment (PISA) and provides important data for educational and social research. PISA reveals the school success of students and handles the factors affecting their performance, as well as allowing comparisons between countries.

In PISA administrations, students are assessed every three years in three subjects: reading, mathematics, and science. Every three years, only one of these areas constitutes the main subject of the application. PISA started with reading literacy as the major domain in 2000 and then continued with the main fields of mathematics and science, respectively. This process has continued in this order until now. In these administrations, cognitive tests are applied to see the extent to which 15-year-old students have the knowledge and skills necessary for participation in societies, while questionnaires are implemented to assess student background factors, school-level factors, and non-cognitive and metacognitive factors. As in previous cycles of PISA, PISA 2018 student questionnaires dealt with non-cognitive and metacognitive variables related to the main subject (reading-related outcomes). In addition to this, it is concerned with non-cognitive variables (dealing with general topics rather than domain-specific topics) called dispositional variables and school-focused variables (learning beliefs and attitudes towards school and achievement goals).

Dispositional Variables in PISA Questionnaires

Dispositional variables are the personality-based contexts that include students' approaches to learning or their avoidance, such as the achievement motives of competitiveness, fear of failure, and work

* Assist. Prof. Dr. Marmara University, Faculty of Education, İstanbul-Türkiye, munevver.rock@gmail.com, ORCID ID: 0000-0003-3572-7982

To cite this article:
Başman,M. (2024). Measurement invariance of achievement motives model: PISA 2018 Turkey sample, *15*(1), 18-34. https://doi.org/10.21031/epod.1302574

Received: 25.05.2023
Accepted: 13.01.2024

mastery; subjective well-being; perseverance; incremental mind-set; and information and communication technology motivation and practices. It is the result of lifelong socialization by parents, teachers, coaches, and one's cultural environment, and shows how behavior gets stronger over time. These variables are important since they are one of the best predictors of achievement and domain-specific outcomes (Organization for Economic Co-operation and Development [OECD], 2019a).

One of the dispositional variables determined in PISA 2018 is achievement motives constructed with work mastery, competitiveness, and fear of failure variables. Henry Murray (1938; as cited in Hangen & Elliot, 2016) introduces the achievement motives and Atkinson (1957) presents a model in which achievement motives are the figures that motivate people to be successful and avoid being unsuccessful in some standards of excellence in certain conditions. Two concepts are mentioned here: the need for achievement and the fear of failure. The need for achievement is introduced with three factors: mastery ("preference for challenging, difficult tasks"), work ("enjoyment of working hard"), and competitiveness ("liking for interpersonal competition and the desire to better others") in the Work and Family Orientation Questionnaire (WOFO) developed by Helmreich and Spence (1978) (Helmreich et al., 1980, p. 4). Then, mastery and work factors are combined as a work mastery motive because mastery and work factors are highly correlated and share important content. This creates a two-dimensional model, work mastery and competitiveness, of the need for achievement (Spence & Helmreich, 1983, as cited in Hangen & Elliot, 2016).

The other factor of the achievement motive is competitiveness. Franken and Brown (1995) state that individuals want to be competitive for different reasons and try to identify these reasons in their study. They develop a scale with five factors ("desire to win, satisfaction that comes from improving one's performance, motivation to put forth effort in competitive situations, satisfaction that comes from performing well, preference for difficult tasks") (Franken & Brown, 1995, p. 178). The first three factors are associated with competitiveness and the last two factors are about work mastery motives, which are the factors in the Questionnaire of Spence and Helmreich mentioned before.

The last variable of the achievement motives is fear of failure. It is defined as "disposition to avoid failure and/or a capacity for experiencing shame and humiliation as a consequence of failure" by Atkinson (1957, p. 360). Shame seems to be an emotional consequence of failure, which is highly disturbing to individuals with a high fear of failure, and has been shown to be associated with avoidance and withdrawal tendencies (Mascolo & Fischer, 1995). In other words, it is a tendency that focuses on avoiding the consequences of failure, unlike the need for achievement (Hangen & Elliot, 2016). The consequences of failure are feared rather than the failure itself (Birney et al., 1969, as cited in Conroy, 2003).

Competitiveness, work mastery, and fear of failure are also defined in the PISA 2018 reports. Competitiveness is stated as a desire to be superior to others. Work mastery is described as a desire to work hard to complete tasks. Fear of failure is expressed as a tendency to avoid potential errors and failures in order to protect themselves. Other PISA cycles assess similar factors but these factors are reviewed and reconstructed in PISA 2018 as the factors of achievement motives. For example, test anxiety was used in the previous PISA cycles, but fear of failure is used instead of test anxiety in PISA 2018. It is stated that fear of failure is a more general tendency to avoid potential mistakes and failures because they are experienced as embarrassing, and this can predict cognitive achievement in real-life situations more than test anxiety (OECD, 2019a).

The Measurement Invariance

The achievement motives can affect students' achievements directly or indirectly. However, when the results obtained from or related to these variables are compared between groups, it is not correct to attribute the differences only to the characteristics of the groups, because these differences between the groups may be due to the measurement tool rather than the characteristics of the groups. It is not certain whether any difference between the groups is because of a true difference or psychometric differences (Cheung & Rensvold, 2002). Differences in scores may be due to many confounding variables, such as familiarity with item response formats, test adaptation, and many other socio-cultural factors. Groups can only be compared when scale scores from different groups measure the same factor of interest on the same metric. Only then can score differences between groups be truly represented and meaningful. Therefore, evidence should be presented to make a factor comparison

across groups (Wu et al., 2007). One of these pieces of evidence is measurement invariance evidence. Drasgow and Kanfer (1985) state that measurement invariance is established when the relationship between observed scores and latent factors is the same across groups and when individuals from different groups having the same scores on the latent factor have the same observed scores. In other words, it means that the probability of an individual's observed score being independent of group membership depends on the true score (Wu et al., 2007).

There are various methods for examining measurement invariance. Khorramdel et al. (2020) indicate that some researchers interested in cross-cultural tradition have given their attention to measurement invariance in non-cognitive measures using the latent variable framework and multigroup confirmatory factor analysis (MGCFA). MGCFA, introduced by Jöreskog (1971), is one of the methods of structural equation models used to determine the measurement invariance. MGCFA examines a large number of issues through a single procedure rather than through many separate procedures. Structural Equation Modeling (SEM) provides direct measurement of how much a measurement model is improved or impaired by various intergroup constraints; this offers a clear advantage over other techniques currently in use (Cheung & Rensvold, 2002). On the other hand, MGCFA has disadvantages and limitations in testing measurement invariance when the number of groups and sample size in the data are large (Ding et al. 2023). Measurement invariance with MGCFA is examined by testing four nested hierarchical models or hypotheses, which are: configural invariance, metric (weak) invariance, scalar (strong) invariance, and strict (residual) invariance (Meredith, 1993; Steenkamp et al., 1998; Vandenberg & Lance, 2000).

Configural invariance is the basic form and the first step of invariance. It is tested whether factors have the same pattern of free and fixed loadings across groups and whether individuals in different groups use the same conceptual framework when answering the scale items (Cheung & Rensvold, 2002; Khojasteh, 2012; Wu et al., 2007; Vandenberg & Lance, 2000). Metric invariance is the equality test for scaling units across groups. It determines whether the item loadings on the factors are the same across groups (Khorramdel et al., 2020). Factor loadings are regression slopes that connect the observed variables to the latent variables of interest and thus represent the expected amount of change in the observed variable for one unit of change in the latent variable (Wu et al., 2007). Scalar invariance is the equality test of the intercepts of the regression equations of the observed scores on the latent variables across groups (Khademi, 2020; Schmitt & Kuljanin, 2008). It is tested whether the mean differences in the observed scores are attributed to the mean differences of the latent variables (Finch & French, 2015; Steinmetz et al., 2009; Tucker et al., 2006). Strict invariance is the equality test of unique variances across groups (Khademi, 2020; Vandenberg & Lance, 2000). It is tested whether the mean or covariance differences in the observed scores are attributed to the mean or covariance differences in the latent variables (Gregorich, 2006; Meade et al., 2006).

The measurement invariance of the questionnaires related to the achievement motives and structures in the PISA application was determined in order to determine the usability of the questionnaires in Turkey. The measurement invariance of various scales in these questionnaires was examined according to some variables such as gender, school type, statistical region, socioeconomic status, and countries. In this study, the measurement invariance of the relevant model is handled according to gender and school type. In PISA applications, the relationship between various information obtained from students through questionnaires and students' literacy performance is examined. In the PISA final reports, the success differences of students in school types and different gender groups and the factors affecting success are discussed in detail (Education Reform Initiative-ERG, 2009; OECD, 2019b). It is a common finding of international and national studies that academic achievement differences between gender and school types have existed for a long time in Turkey (Berberoğlu & Kalender, 2005; Suna et al., 2020). In comparisons of questionnaires and tests by gender and school type, it is assumed that the measurements are equally valid in different groups, and measurement invariance can be ignored. The studies conducted in Turkey in the last 10 years examining the measurement invariance of the relevant structure according to gender and school type are given below.

Researchers have examined the invariance of the scales or models in the PISA survey according to gender, school type, countries, statistical region, socioeconomic status, and years (Ardıç & Gelbal , 2017; Başusta & Gelbal, 2015; Demir, 2016; Gülleroğlu, 2017; Güngör & Atalay Kabasakal, 2020; İmrol, 2017; Kıbrıslıoğlu, 2015; Kıbrıslıoğlu Uysal & Akın Arıkan, 2018; Uyar & Doğan, 2014; Uyar

& Kaya Uyanık, 2019). It has been observed that some models provide full measurement invariance (measurement invariance in all four steps is supported) according to the relevant variables, while others do not. For example, while Başusta and Gelbal (2015), Kıbrıslıoğlu (2015), Gülleroğlu (2017), Kıbrıslıoğlu Uysal and Akın Arıkan (2018), and Güngör and Kabasakal (2020) found full measurement invariance according to gender in the models established in their studies, Demir (2016), Ardıç and Gelbal (2017), and Uyar and Kaya Uyanık (2019) show that full measurement invariance has not been established according to gender.

Examining research conducted outside of Turkey, some studies (Adsul & Kamble, 2008; Awan et al., 2011; Nien & Duda, 2008; Shekhar & Devi, 2012; Tang & Lu, 2013) have demonstrated full measurement invariance across gender, while others (Freund et al., 2011; Karaman & Smith, 2019) have not. There are also studies examining differences in attitudes towards competition, motivation to master tasks, and fear of failure according to gender, and they found the full measurement invariance of the scales considering gender because the results of group differences obtained without measurement invariance are questionable (De Paola et al., 2015; Eber et al., 2021; Givord, 2020; OECD, 2019b; Severiens & ten Dam, 1998).

Whether it is PISA applications or other international applications, the results of these applications guide the development of education policies. In order for the results of the applications to be meaningful and valid, the measurement invariance of the measurement tools (achievement tests and questionnaires) used in the research should be ensured between subgroups such as gender, socio-economic level, school type, and culture; otherwise the comparisons will not be meaningful and valid (Vandenberg & Lance, 2000). It was observed that sufficient measurement invariance studies were not conducted on the student questionnaires of the PISA 2018 application. For this reason, examining the measurement invariance for the achievement motives model used in the PISA 2018 application will provide evidence for the validity of the model and determining whether the group comparisons are meaningful according to the scores obtained will contribute to a more accurate interpretation of the results. Thus, it is thought that examining the measurement invariance of the model, which has not yet been made in the literature, will fill the gap in the field. In Turkey, the main subject of PISA applications can be examined in terms of affective variables, and affective variables can be handled in terms of demographic variables such as gender and school type. Achievement differences between school types and gender in Turkey can be relatively high (Berberoğlu & Kalender, 2005; Suna et al., 2020). Examining the measurement invariance of the achievement motivation model in terms of gender and school type is important for Turkey in the context of equality in education. For these reasons, the aim of this study is to examine the measurement invariance of the achievement motives model constructed by attitudes towards competition, motivation to master tasks, and fear of failure scales in the PISA 2018 student questionnaire with regard to gender and school type in the Turkey sample. Answers were sought for the following questions in the study:

(1) What are the levels of fit of the achievement motives model with the data obtained from the whole group, gender, and school type subgroups?

(2) Does the achievement motives model hold measurement invariance across gender and school type subgroups?

Methods

Research Design

In this study, it is examined whether measurement invariance of the achievement motives model, including attitudes towards competition, motivation to master tasks and fear of failure scales, is held across gender and school type in the PISA 2018 application in the Turkey sample. This study is descriptive research and aims to determine an existing situation concerning the psychometric characteristics of the measurements obtained from the scales (Fraenkel et al., 2012; Karasar, 2019).

Population and Sample

In the PISA 2018 application, 38 OECD member countries and 41 non-member countries participated. There are 600,000 students, representing about 32 million in total (OECD, 2019b). In this research, the measurement invariance of the achievement model is examined in the Turkey sample. Turkey participated in the PISA 2018 application with 6890 students from 186 schools, representing approximately 884,971 students at the age of 15. Schools in determining the Turkey sample of the PISA 2018 application school type, Regional Units for Statistics Classification Level 1, administrative

form of the school, location of the school, and gender distribution layers were used. After the schools were determined, the students who would participate in the application at the selected schools were randomly selected (MEB, 2019). The Turkish sample consists of 6442 students. Table 1 presents the distribution of the students in the study groups according to their genders and school types.

Table 1.
Distribution of the Students in Study Group according to Gender and School Types

School	Gender				Total	
	Female		Male			
	n	%	*n*	%	*n*	%
Anatolian High School	1456	51.2	1386	48.8	2842	44.1
Vocational and Technical Anatolian High School	881	44.0	1122	56.0	2003	31.1
Anatolian Imam and Preacher High School	469	54.1	398	45.9	867	13.5
Science, Social Sciences, Multi-Programme Anatolian, Anatolian Sport/Anatolian Fine Arts High School	417	57.1	313	42.9	730	11.3
Total	3223	50.0	3219	50.0	6442	100

This study is carried out with 3223 (about 50%) female and 3219 (about 50%) male students. The schools attended by these students are Anatolian High School (44.1%), Vocational and Technical Anatolian High School (31.1%), Anatolian Imam and Preacher High School (13.5%), and Science, Social Sciences, Multi-Programme Anatolian, and Anatolian Sport/Anatolian Fine Arts High School (11.3%).

Data Collection

Data obtained from the PISA 2018 student questionnaire is used in this study. The data file for PISA 2018 can be found at the OECD PISA website, https://www.oecd.org/pisa/data/2018database/. Within the scope of the research, achievement motives model composed of attitudes towards competition (ST181), motivation to master tasks (ST182), and fear of failure (ST183) scales from the student questionnaire are used in 4-point Likert-type scales such as "strongly disagree (1), disagree (2), agree (3), and strongly agree (4)". While attitudes towards competition and fear of failure scales consist of three items, motivation to master tasks consists of four items (OECD, 2019b). The items of the achievement motives model constructed with these scales, as mentioned in PISA 2018 reports (OECD, 2019a) are shown in Table 2.

Table 2.
The Achievement Motives Model Items

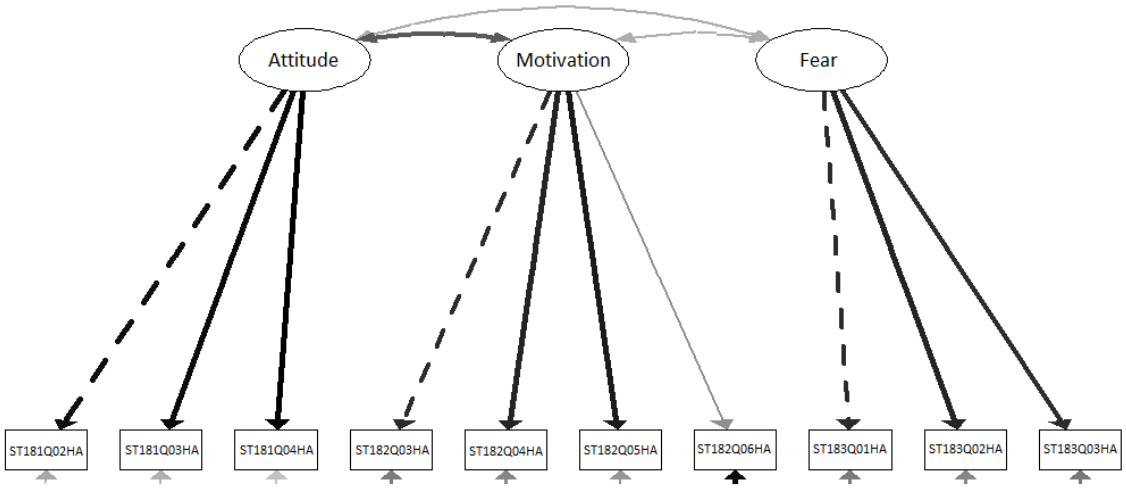
Code of Items	Items
ST181	Attitudes towards competition items
ST181Q02HA	I enjoy working in situations involving competition with others.
ST181Q03HA	It is important for me to perform better than other people on a task.
ST181Q04HA	I try harder when I'm in competition with other people.
ST182	Motivation to master tasks items
ST182Q03HA	I find satisfaction in working as hard as I can.
ST182Q04HA	Once I start a task, I persist until it is finished.
ST182Q05HA	Part of the enjoyment I get from doing things is when I improve on my past performance.
ST182Q06HA	If I am not good at something, I would rather keep struggling to master it than move on to something I may be good at.
ST183	Fear of failure items
ST183Q01HA	When I am failing, I worry about what others think of me.
ST183Q02HA	When I am failing, I am afraid that I might not have enough talent.
ST183Q03HA	When I am failing, this makes me doubt my plans for the future.

Data Analysis

The data were primarily organized and examined to see whether they met the assumptions of the structural equation modeling analysis. The arrangement of the data and the control of the assumptions were made with IBM SPSS Statistics (Version 26). Missing data, outlier values, sample size, multicollinearity, and linearity were examined (Kline, 2015; Tabachnick & Fidell, 2013). Firstly, lower-secondary school data is excluded because of the very limited number of observations (n=22). Remaining cases with missing data were also considered inconsequential because the missing data rate is less than 2% and the missing data is missing completely at random according to the MCAR test (p>.05). Therefore, the listwise method was used (Acuna & Rodriguez, 2004; Kline, 2015, Nakagawa, 2015; Schafer, 1999). Multivariate outliers were computed from the Mahalanobis distance and 129 values were found to show multivariate outliers (p<.001). By excluding individuals with these values from the dataset, analysis was continued with 6442 individuals. It was seen that the dataset obtained from 6442 individuals met the sample size, multicollinearity (examining the variance inflation factor, condition index and tolerance values), and linearity (using scatter plot) and was suitable for SEM analysis. After checking the assumptions, the data were analyzed and the measurement model was established. In this research, measurement invariance of the achievement motives model was examined by MGCFA. The three-factor model analyzed in this study is shown in Figure 1.

Figure 1.

The Achievement Motives Measurement Model



The achievement motives model in which variables were in the specified dimensions was established and it was tested with confirmatory factor analysis (CFA) using SEM to analyze the compatibility of this model with the dataset. The structural equation model was applied with the lavaan package (Rosseel et al., 2022) in R software package (Version 4.0.2). The fit between the model and the data was examined with the goodness of fit statistics. Even though there were several parameter estimation methods for ordinal variables used in CFA/SEM analysis, in this research, the Unweighted Least Squares (ULS) method was used for estimations. The reasons to choose the ULS method include: it is one of the most common methods used for ordinal variables and gives more accurate parameter estimations than diagonally weighted least squares (DWLS) and Maximum likelihood (ML) methods (Forero et al., 2009; Koğar & Yılmaz Koğar, 2015; Yang-Wallentin et al., 2010).

According to the results of CFA, MGCFA was used to determine whether the variables show measurement invariance across gender and school type. Although there are various methods (e.g. alignment method, Bayesian structural equation models) in the examination of measurement invariance, the reason for choosing MGCFA is that MGCFA examines the equivalence of covariance structures, works with latent variables instead of observed variables, and latent means analysis is more sensitive than traditional statistical methods to detect between-group differences (Sehee et al., 2003; Vandenberg & Lance, 2000). Measurement invariance was examined by MGCFA including a sequence of tests of four nested hierarchical models or hypotheses, which are: configural invariance, metric (weak) invariance, scalar (strong) invariance, and strict (residual) invariance (Meredith, 1993;

Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). The fit indices of the hierarchically obtained models were examined. While evaluating the fit between the model and the data, the values of chi-square (χ^2), the root mean squared error of approximation (RMSEA; Steiger, 1989), the standardized root mean square residual (SRMR; Bentler, 1995), the comparative fit index (CFI; Bentler, 1990), Tucker–Lewis Index (TLI; Tucker & Lewis, 1973), and the Relative Centrality Index (RNI; McDonald & Marsh, 1990) were taken into consideration.

The value χ^2 is a function of the sample size and tends to reject the null hypothesis when the sample size is large. In other words, the χ^2 test may reject insignificant model-data differences and it is not sufficient by itself (Wu et al., 2007). The deviations of the variables from the normal distribution can inflate goodness-of-fit test statistics (Finney & DiStefano, 2006; Kaplan, 2000). It is thought that it would not be sufficient by itself because the χ^2 is sensitive to sample size and model complexity. For these reasons, Vandenberg and Lance (2000) recommend four indices (RMSEA, SRMR, TLI, and RNI) for overall model fit. RMSEA, SRMR, TLI and RNI are sensitive to misspecified models. SRMR is particularly sensitive to factor covariance misspecification, while others are sensitive to factor loading misspecification. In addition, TLI and RNI are independent of sample size. The reference values for fit indices are stated as follows: .05<RMSEA≤.08 is an acceptable fit, RMSEA≤ .05 is a good fit; .05<SRMR≤.08 is an acceptable fit, SRMR≤.05 is a good fit; .90≤CFI<.95 is an acceptable fit, CFI≥.95 is a good fit; .90≤TLI<.95 is an acceptable fit, TLI≥ .95 is a good fit; .90≤RNI<.95 is an acceptable fit, RNI≥ .95 is a good fit (Hooper et al., 2008; Hu & Bentler, 1999; Tabachnick & Fidell, 2013; Vandenberg & Lance, 2000).

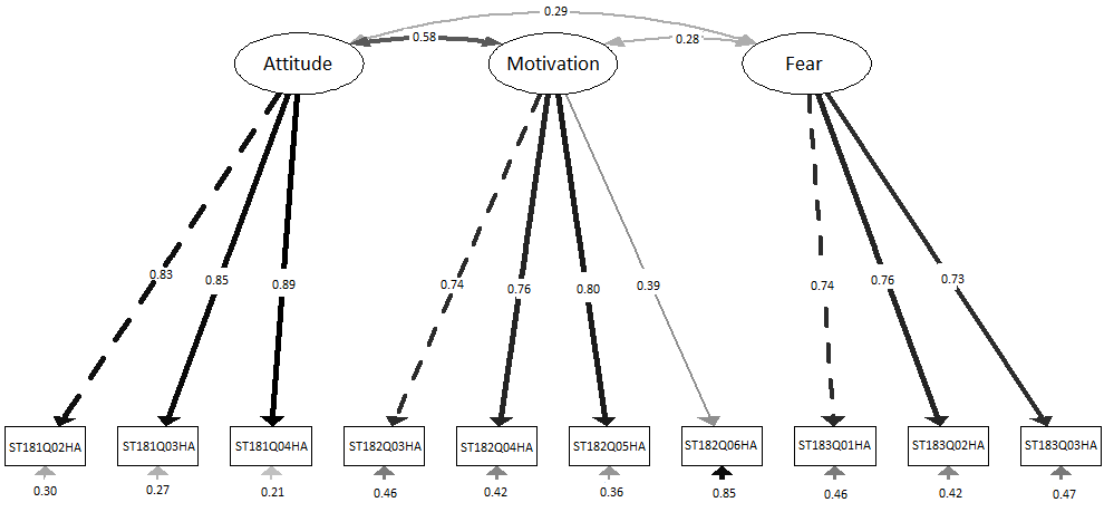
Cheung and Rensvold (2002) state that the likelihood-ratio (LR) test (the chi-square difference test-$\Delta\chi^2$) is generally used to determine model fit differences but $\Delta\chi^2$ test is sensitive to sample size and model complexity as χ^2 test. Yandı et al. (2017) stated that $\Delta\chi^2$ are affected by the degree of freedom and sample size. Dimitrov (2010) indicates that some researchers (e.g., Cheung & Rensvold, 2002; Little, 1997; Vandenberg & Lance, 2000) suggested using changes in other fit statistics to test for measurement invariance because $\Delta\chi^2$ is sensitive to sample size. Şekercioğlu (2018) also agrees that χ^2 is not a practical test for model fit because of statistically sensitive test for large samples and he recommends the use of the most frequently used alternative comparative fit indices like CFI, TLI, and RMSEA instead of χ^2. Cheung and Rensvold (2002) suggest the use of ΔCFI, ΔGamma hat, and ΔMcDonald's Noncentrality Index (ΔMcDonald's NCI) values, which are independent of model parameters and sample size. Furthermore, they indicate the cut-off values as ΔCFI ≤-.01, ΔGamma hat≤-.001, and ΔMcDonald's NCI≤-.02, which means the null hypothesis of invariance should not be rejected. However, Strijbos et al. (2021) state that there is no consensus for the cutoff value for ΔGamma hat and Meade et al. (2006) also state that the value of -.001 may be overly strict because it is affected by small differences in factor loadings. For these reasons, in this study, fit indices (χ^2, RMSEA, CFI, TLI, SRMR, and RNI) of models in addition to the differences of CFI, Gamma hat, and McDonald's NCI values between models are examined to determine measurement invariance. The measurement invariance is tested with the lavaan package (Rosseel et al., 2022) in R software package (Version 4.0.2).

Results

Results on Testing of the Measurement Model

The data were primarily organized and examined to see whether they met the assumptions of SEM analysis as mentioned data analysis (missing data, outlier values, sample size, multicollinearity, and linearity). After checking the assumptions, the three-factor model was established and it was tested with CFA using SEM to analyse the compatibility of this model with the dataset. The model and coefficients obtained according to the results of CFA are given in Figure 2.

Figure 2.
The Achievement Motives Model Path Diagram



The model data fit for the model and subgroups was examined by referring to the indices indicated in Table 3.

Table 3.
Fit Indices of the Achievement Motives Model and Subgroups

Groups	χ^2 (df)	χ^2/df	RMSEA	SRMR	CFI	TLI	RNI
Achievement Motives Model (complete data)	393.129 (32)	**12.285**	.042	.039	.985	.979	.985
Female	161.988 (32)	**5.062**	.036	.037	.986	.980	.986
Male	226.425 (32)	**7.076**	.043	.040	.987	.982	.987
Anatolian High School	187.584 (32)	**5.862**	.041	.041	.982	.975	.982
Vocational and Technical Anatolian High School	123.515 (32)	3.860	.038	.038	.991	.987	.991
Anatolian Imam and Preacher High School	35.207 (32)	3.180	.011	.033	.999	.999	.999
Science, Social Sciences, Multi-Programme Anatolian, Anatolian Sport/Anatolian Fine Arts High School	59.596 (32)	3.930	.034	.046	.989	.985	.989

Note. χ^2 = Chi-square; RMSEA = Root Mean Squared Error of Approximation; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RNI = Relative Centrality Index; Rejections of model invariance were highlighted in bold, **p*<.05.

When the goodness of fit statistics of the scale scores are examined, it is seen that the obtained values show good fits, except χ^2 (χ^2/df = 12.285, *p*<.05; RMSEA = .042, SRMR = .039, CFI = .985, TFI = .979, RNI=.985). This situation can be explained by the sample size and model complexity sensitivity of χ^2. As a result of the analysis, it was determined that the model was compatible with the data because the other fit indices were within acceptable limits. Furthermore, it was seen that the fit between the model and the data across groups was provided (RMSEA≤.05, SRMR≤.05, CFI ≥.95, TFI ≥.95, RNI≥.95).

Results on Testing of the Measurement Invariance by Gender

The measurement invariance of the achievement motives model, which includes three scales, was examined by testing four nested hierarchical models, which are configural invariance, metric invariance, scalar invariance, and strict invariance. Multigroup CFA findings for the three-factor structure equality of the achievement motives model are given in Table 4 according to gender.

Table 4.

Fit Indices for Invariance Tests by Gender Groups

Model	χ2 (df)	RMSEA	SRMR	CFI	TLI	RNI
Configural	388.412 (64)	.040	.036	.987	.982	.987
Metric	438.086 (71)	.040	.038	.985	.981	.985
Scalar	500.368 (78)	.041	.041	.983	.980	.983
Strict	521.799 (88)	.039	.042	.982	.982	.982

Model	Δχ2 (Δdf)	ΔCFI	ΔGamma hat	ΔMcDonald's NCI
Configural	-	-	-	-
Metric	**49.674* (7)**	-.002	-.001	-.003
Scalar	**62.282* (7)**	-.002	**-.002**	-.004
Strict	**21.431* (10)**	-.001	.000	.000

Note. χ^2 = Chi-square; RMSEA = Root Mean Squared Error of Approximation; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RNI = Relative Centrality Index; McDonald's NCI = McDonald's Noncentrality Index; Δ… = Change in fit index. Rejections of model invariance are highlighted in bold, **p*<.05.

The measurement invariance of the achievement model of the PISA 2018 student questionnaire, as shown in Table 4, was established according to gender in the Turkey sample. The four nested hierarchical models were examined and it was seen that the values of the fit indices were acceptable, except χ^2. The differences of χ^2 were significant (p<.05), but it was stated that they should not be evaluated alone because χ^2 is sensitive to the sample size and model complexity in confirmatory factor analytic tests of measurement invariance (Meade et al., 2006). In practice, chi-square is not considered to be a very useful fit index by most researchers because it is affected by several factors (Newsom, 2020). LR tests reject the null hypothesis with too much power if the sample size is large, as the case in our study. In other words, LR tests may reject trivial model-data differences and thus lose practical usefulness (Wu et al., 2007). As the sample size increases, the chi-square value increases, leading to the problem that plausible models are rejected due to trivial discrepancies in measurement invariance tests (Khojasteh, 2012; Wang, 2008; Chen, 2007; Brannick, 1995). Since $\Delta\chi^2$ test is sensitive to sample size, and the sample size in our study (6442) is very high, using the differences in other fit statistics is suggested by the researchers (Cheung & Rensvold, 2002; Şekercioğlu, 2018; Vandenberg & Lance, 2000) to test for measurement invariance. Thus, the goodness of fit indices and ΔCFI, ΔGamma hat, and ΔMcDonald's NCI values were taken as basis in line with the findings of the MGCFA. First, the configural invariance step was provided considering fit indices (RMSEA = .040, SRMR = .036, CFI = .987, TLI = .982, RNI=.987) because the fit indices had acceptable values, except χ^2 ($\chi^2_{(64)}$= 388.412, p<.05). In other words, individuals in different gender groups use the same conceptual framework when answering the scale items. The metric invariance was tested in the second step, and it was observed that metric invariance was held according to fit indices (except χ^2), and the differences of CFI and McDonald's NCI (RMSEA = .040, SRMR = .038, CFI = .985, TLI = .981, RNI=.985; ΔCFI = -0.002, ΔMcDonald's NCI=-.003). On the other hand, LR tests and ΔGamma hat showed that there was no metric invariance ($\chi^2_{(71)}$= 438.086, p<.05; $\Delta\chi^2_7$=49.674, p<.05; ΔGamma hat = -.0013). It was observed that different tests provided different results according to metric invariance. However, as stated above, since the study included a large sample size, LR tests may not provide reliable results due to their sensitivity to sample size. For the ΔGamma hat test, the exact value of -0.0013 was only slightly out of the acceptable range, and it is noted that there is no consensus for the cutoff value for ΔGamma hat (Strijbos et al., 2021). Meade et al. (2006) also state that the value of -

.001 may be overly strict because it is affected by small differences in factor loadings. As a result of these discussions, even though metric invariance was not held based on ΔGamma hat and LR tests, since it was held according to most of the fit indices, ΔCFI and ΔMcDonald's NCI, it was concluded that the factor loadings of the model were the same for male and female groups as in the factor structures of the model. The next step was to check scalar invariance after the configural and metric invariance were found to be satisfied. When the scalar invariance was examined, similar to metric invariance, it was seen that scalar invariance was held according to fit indices (except χ^2), and the differences of CFI and McDonald's NCI (RMSEA = .041, SRMR = .041, CFI = .983, TLI = .980, RNI=.983; ΔCFI = -0.002, ΔMcDonald's NCI=-.004). On the other hand, LR tests and ΔGamma hat showed that there was no scalar invariance ($\chi^2_{(78)}$= 500.368, p<.05; $\Delta\chi^2_7$=62.282, p<.05; ΔGamma hat = -.0017). Due to similar reasons as stated above for the metric invariance, even though scalar invariance was not held based on ΔGamma hat and LR tests, since it was held according to most of the fit indices, ΔCFI and ΔMcDonald's NCI, it was concluded that the regression constants were the same for male and female groups. In the last step, it was observed that the strict invariance was held according to fit indices (except χ^2), and the differences of CFI, Gamma hat and McDonald's NCI (RMSEA = .039, SRMR = .042, CFI = .982, TLI = .982, RNI=.982; ΔCFI = -0.001, ΔGamma hat = -.000, ΔMcDonald's NCI=-.000). On the other hand, LR tests showed that there was no strict invariance ($\chi^2_{(88)}$= 521.799, p<.05; $\Delta\chi^2_{10}$=21.431, p<.05). Thus, because of the same reasons as stated above for the LR tests, since most of the tests agree to have strict invariance, it was concluded that the residual variances for each item are the same in addition to equal factor loadings, slopes and intercepts across groups. Considering the results of the majority of the tests, the full measurement invariance of the achievement motives model is accepted to be held by gender subgroups. As a result, all comparisons made for gender regarding the model will be meaningful according to these findings.

Results on Testing of the Measurement Invariance by School type

Multigroup CFA findings for the three-factor structure equality of the achievement motives model are given in Table 5 according to school type.

Table 5.

Fit Indices for Invariance Tests by School Type Groups

Model	χ2 (df)	RMSEA	SRMR	CFI	TLI	RNI
Configural	405.902 (128)	.037	.037	.989	.984	.989
Metric	443.441 (149)	.035	.038	.988	.986	.988
Scalar	517.921 (170)	.036	.041	.986	.985	.986
Strict	546.103 (200)	.033	.042	.986	.987	.986

Model	Δχ2 (Δdf)	ΔCFI	ΔGamma hat	ΔMcDonald's NCI
Configural	-	-	-	-
Metric	**37.539* (21)**	-.001	.000	-.002
Scalar	**74.480* (21)**	-.002	**-.002**	-.004
Strict	**28.182* (30)**	.000	.000	.000

Note. χ^2 = Chi-square; RMSEA = Root Mean Squared Error of Approximation; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RNI = Relative Centrality Index; McDonald's NCI = McDonald's Noncentrality Index; Δ… = Change in fit index. Rejections of model invariance are highlighted in bold, *p<.05.

The measurement invariance of the achievement model of the PISA 2018 student questionnaire, as shown in Table 5, was established according to school type in the Turkey sample. As mentioned before, the differences of χ^2 were significant (p<.05), but as stated above they should not be evaluated alone because χ^2 is sensitive to the sample size and model complexity, so the fit indices and ΔCFI, ΔGamma hat, and ΔMcDonald's NCI values were also examined. First, the configural invariance step was provided considering fit indices (RMSEA = .037, SRMR = .037, CFI = .989, TLI = .984, RNI = .989) because the fit indices had acceptable values, except χ^2 ($\chi^2_{(128)}$= 405.902, p<.05). In other words, individuals in different school type groups use the same conceptual framework when answering the

scale items. Then, in the second step, the metric invariance was tested, and it was observed that metric invariance was held according to fit indices (except χ^2) (RMSEA = .035, SRMR = .038, CFI = .988, TLI = .986 RNI = .988), and the differences of CFI, Gamma hat and McDonald's NCI (ΔCFI = -0.001, ΔGamma hat = -.000, ΔMcDonald's NCI=-.002). On the other hand, LR tests showed that there was no metric invariance ($\chi^2_{(149)}$= 443.441, p<.05; $\Delta\chi^2_{21}$=37.539, p<.05). As stated above, since the study included a large sample size, LR tests may not provide reliable results due to their sensitivity to sample size. Even though metric invariance was not held based on LR tests, since it was held according to most of the fit indices, ΔCFI, Δ Gamma hat and ΔMcDonald's NCI, it was concluded that the factor loadings of the model were accepted to be the same for school type groups as in the factor structures of the model. The next step was to check scalar invariance after the configural and metric invariances were found to be satisfied. When the scalar invariance was examined, it was seen that scalar invariance was held according to fit indices (except χ^2), and the differences of CFI and McDonald's NCI (RMSEA = .036, SRMR = .041, CFI = .986, TLI = .985, RNI=.986, Gamma hat = .975; ΔCFI = -0.002, ΔMcDonald's NCI=-.004). On the other hand, LR tests and ΔGamma hat showed that there was no scalar invariance ($\chi^2_{(170)}$= 517.921, p<.05; $\Delta\chi^2_{21}$=74.480, p<.05; ΔGamma hat = -.0017). Due to similar reasons as stated above for the scalar invariance considering gender, even though scalar invariance was not held based on ΔGamma hat and LR tests, since it was held according to most of the fit indices, ΔCFI and ΔMcDonald's NCI, it was concluded that the regression constants were the same for school type groups. In the last step, it was observed that the strict invariance was held according to fit indices (except χ^2), and the differences of CFI, Gamma hat and McDonald's NCI (RMSEA = .033, SRMR = .042, CFI = .986, TLI = .987, RNI=.986; ΔCFI = -.000, ΔGamma hat = -.000, ΔMcDonald's NCI=-.000). On the other hand, LR tests showed that there was no strict invariance ($\chi^2_{(200)}$= 546.103, p<.05; $\Delta\chi^2_{30}$=28.182, p<.05). Thus, because of the same reasons as stated above for the LR tests, since most of the tests agree to have strict invariance, it was concluded that the residual variances for each item are the same in addition to equal factor loadings, slopes, and intercepts across groups. Considering the results of the majority of the tests, the full measurement invariance of the achievement motives model is accepted to be held by school type subgroups. As a result, the full measurement invariance of the achievement motives model held by school type subgroups. All comparisons made for school type regarding the model will be meaningful according to these findings.

Discussion

The importance of the individual's affective characteristics in acquiring behaviors and skills in the cognitive domain is known. Affective characteristics also affect school success. Given the role that affective learning outcomes play in shaping students' future behavior, educators should pay attention to students' affective characteristics. Lessons should be developed by taking into account the three learning domains of education, namely cognitive, psychomotor, and affective, and these three domains should be included in the education process. The level of acquisition of these knowledge, skills, and affective characteristics should also be measured and education policies should be planned accordingly. Before making measurements in the affective field, the measurement invariance of the measurement tools to be used must be demonstrated. In this way, it can be determined whether the results obtained are due to the measurement tool or not.

For these reasons, while observing cognitive skills, the individual's affective characteristics should also be taken into account. In international assessment administrations, besides measuring knowledge and skills in cognitive fields, it is also aimed to measure affective characteristics. Some of affective characteristics measured in PISA 2018 are the attitudes towards competition, motivation to master tasks, and fear of failure scales, which are under the achievement motives model.

When it is desired to examine the affective characteristics of individuals or to carry out studies related to these characteristics, first of all, the measurement invariance of the measurement tools that measure these characteristics should be ensured. Measurement invariance is important as it can provide evidence about whether tests/questionnaires measure the same factor in the same way in different groups. In this research, the measurement invariance of the achievement motives model was examined according to gender and school type in the PISA 2018 application in the Turkey sample. The achievement motives model consists of the attitudes towards competition, motivation to master tasks,

and fear of failure scales in the PISA 2018 administration. The three-factor model, established for achievement motives, was tested for the complete datasets, as well as for each gender group and school type group. Confirmatory factor analysis results show that the goodness of fit indices of the measurement model are at acceptable levels except for the lower-secondary school group. Thus, the data of the lower-secondary school group were excluded from the school type dataset. The measurement invariance of the achievement motives model was examined according to gender and school type groups via Multigroup Confirmatory Factor Analysis based on four models. According to MGCFA results, the full measurement invariance of the achievement motives model is accepted to be held by gender and school type subgroups because the values of fit indices and their change are acceptable values, except $\chi 2$. It is noted that both $\Delta\chi 2$ are sensitive to sample size and model complexity (Cheung & Rensvold, 2002; Dimitrov, 2010; Şekercioğlu, 2018; Yandı et al., 2017), and there is no consensus for the cutoff value for ΔGamma hat (Strijbos et al., 2021); the value of -.001 may be overly strict because it is affected by small differences in factor loadings (Meade et al., 2006). Because of these reasons, even though measurement invariance is not held according to LR test results and scalar invariance was not held based on ΔGamma hat, since most of the fit indices, ΔCFI and ΔMcDonald's NCI test results indicate measurement invariance, considering the results of the majority of the tests, the full measurement invariance is accepted to be held according to gender and school type.

Gender differences in achievement motives have been examined in various studies, and there are studies that found differences in achievement motives according to gender (Adsul & Kamble, 2008; Awan et al., 2011; Shekhar & Devi, 2012) as well as studies that do not find any difference (Khan et al., 2011; Yeung et al., 2012; Kaura & Sharma, 2015). In addition, there are some studies examining gender differences in attitudes towards competition, motivation to master tasks, and fear of failure (eber et al.,2015; Eber et al., 2021; Givord, 2020; OECD, 2019b; Severiens & ten Dam, 1998). Before examining the differences by gender, the measurement invariance of achievement motives should be examined. Otherwise, it cannot be determined whether the differences obtained are due to the measurement tool or due to the real differences. The measurement invariance of achievement motives used in the aforementioned studies was examined across genders. Nien and Duda (2008) and Tang and Lu (2013) found that the full measurement invariance held across genders. It can be said that these findings are in parallel with this research. On the other hand, Freund et al. (2011) and Karaman and Smith (2019) found the full measurement invariance is not established across genders.

When the studies examining the achievement motives and their related factors across gender and school type in PISA applications in Turkey are examined, it can be said that these findings are in parallel with the studies by Başusta and Gelbal (2015), Kıbrıslıoğlu (2015), Gülleroğlu (2017), Kıbrıslıoğlu Uysal and Akın Arıkan (2018), and Güngör and Kabasakal (2020) in terms of showing the full measurement invariance of the models according to gender. On the other hand, the studies of Demir (2016), Ardıç and Gelbal (2017), and Uyar and Kaya Uyanık (2019) state that the full measurement invariance is not established according to gender.

Due to the relatively high differences in achievement between school types in Turkey, the results related to the achievement motives model obtained without considering the measurement invariance in the school type may not be valid and reliable (Berberoğlu & Kalender, 2005; Suna et al., 2020). Comparisons by gender or school type will not be meaningful if full measurement invariance is not provided. In this study, the achievement motives model shows full measurement invariance by gender and school type. It can be said that these findings are in parallel with the study by Ardıç and Gelbal (2017), İmrol (2017) in terms of showing the full measurement invariance of the models according to school type, while the study of Uyar and Doğan (2014) does not establish the full measurement invariance according to school type. These results suggest that gender and school type-related measurement invariance merits attention in achievement motives research.

The results of the measurement invariance carried out in this study show that the psychometric qualities of the measurements obtained from the measurement model, which consists of items in the PISA student questionnaire that aim to reveal students' attitudes towards competition, motivation to master tasks, and fear of failure, can be generalized among gender and school type groups. It can be said that the difference between the groups is not due to the measurement tool. The measurements obtained from the achievement motives model items could be generalized among the school groups

and gender, and provide reliable and valid measurements for determining the achievement motives of the students. In this regard, the scores obtained from the achievement model can be used in comparisons according to gender or school type. The results obtained from the competitive attitudes, motivation to master the task, and fear of failure scale can be used reliably and validly to examine the differences between individuals considering gender and school type variables. In addition, the researcher is advised to be cautious when comparing scores in dispositional variables of different groups if there is no evidence about measurement invariance.

This study is limited to the responses given to the achievement motives model in the PISA 2018 student questionnaire towards attitudes towards competition, motivation to master tasks, and fear of failure scale items. In addition, the measurement invariance of the achievement motives model is limited for the group of students at the age of 15 in Turkey. If the achievement motives model is to be used in different age groups, first of all, measurement invariance should be satisfied for that age group, and then the achievement motives model and their scales should be used. The scales or the established models consisting of the scales from PISA should be examined to obtain measurement invariance evidence across groups before using them for the purpose of comparing groups and generalizing the findings. In future studies, researchers can repeat the research using other groups and different models or scales. In addition, measurement invariance studies of the same models can be conducted in different countries. The widely used MGCFA method is used in this study. Other methods can be used and compared to examine measurement invariance because MGCFA has limitations in testing measurement invariance when the number of groups and sample size are large.

Declarations

Conflict of Interest: The author reports there are no competing interests to declare.

Ethical Approval: I declare that all ethical guidelines for the author have been followed. This study does not require any ethics committee approval as it includes open-access data.

References

Acuna, E., & Rodriguez, C. (2004). The treatment of missing values and its effect on classifier accuracy. In *Classification, clustering, and data mining applications* (pp. 639-647). Springer, Berlin, Heidelberg.

Adsul, R. K., Kamble, V., & Sangli, K. W. (2008). Achievement motivation as a function of gender, economic background and caste differences in college students. *Journal of the Indian Academy of Applied Psychology*, *34*(2), 323-327.

Ardıç, E., & Gelbal, S. (2017). Cross-group equivalence of interest and motivation items in PISA 2012 Turkey sample. *Eurasian Journal of Educational Research, 17*(68), 221-238. http://dx.doi.org/10.14689/ejer.2017.68.12

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6, Pt.1), 359–372. https://doi.org/10.1037/h0043445

Awan, R. U. N., Noureen, G., & Naz, A. (2011). A Study of Relationship between Achievement Motivation, Self Concept and Achievement in English and Mathematics at Secondary Level. *International education studies*, *4*(3), 72-79. http://dx.doi.org/10.5539/ies.v4n3p72

Başusta, N. B & Gelbal, S. (2015). Gruplar arası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacetepe Üniversitesi Eğitim Fakültesi Dergisi*, *30*(4), 80-90.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). *EQS structural equations program manual*. Multivariate Software.

Berberoğlu, G., & Kalender, İ. (2005). Öğrenci başarısının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi. *Journal of Educational Sciences & Practices*, *4*(7).

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of organizational behavior, 16*, 201–213.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural equation modeling: A multidisciplinary journal*, *19*(3), 372-398. http://dx.doi.org/10.1080/10705511.2012.687671

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Conroy, D. E. (2003). Representational models associated with fear of failure in adolescents and young adults. *Journal of Personality*, *71*(5), 757-783. https://doi.org/10.1111/1467-6494.7105003

De Paola, M., Ponzo, M., & Scoppa, V. (2015). Gender differences in attitudes towards competition: Evidence from the Italian scientific qualification. IZA ddiscussion paper. No: 8859.

Demir, E. (2016). Testing measurement invariance of the students' affective characteristics model across gender sub-groups. *Kuram ve Uygulamada Eğitim Bilimleri, 17*(1), 47-62. https://doi.org/10.12738/estp.2017.1.0223

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121-149. https://doi.org/10.1177/0748175610373459

Ding, Y., Yang Hansen, K., & Klapp, A. (2023). Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFA and the alignment method. *European Journal of Psychology of Education*, *38*(2), 709-732.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, *70*(4), 662–680. https://doi.org/10.1037/0021-9010.70.4.662

Eber, N., François, A., & Weill, L. (2021). Gender, age, and attitude toward competition. *Journal of Economic Behavior & Organization*, *192*, 668-690. https://doi.org/ 10.1016/j.jebo.2021.10.022

Education Reform Initative (Eğitimde Reform Girişimi) (2009). *Eğitimde eşitlik: Politika analizi ve öneriler.* ERG Raporları. Retrieved from https://www.egitimreformugirisimi.org/wp-content/uploads/2017/03/Egitimde_Esitlik_Politika_Analizi_ve_Oneriler_1.pdf.

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. Routledge.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock and R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269- 314). Greenwich, CT: Information Age.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625 – 641. https://doi.org/ 10.1080/10705510903203573

Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Franken, R. E., & Brown, D. J. (1995). Why do people like competition? The motivation for winning, putting forth effort, improving one's performance, performing well, being instrumental and expressing forceful/aggressive behavior, *Personality and Individual Differences, 19*(2), 175-184. https://doi.org/10.1016/0191-8869(95)00035-5

Freund, P. A., Kuhn, J. T., & Holling, H. (2011). Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, *51*(5), 629-634. https://doi.org/ 10.1016/j.paid.2011.05.033

Givord, P. (2020), "Do boys and girls have similar attitudes towards competition and failure?", *PISA in Focus*, No. 105, OECD Publishing, Paris. https://doi.org/10.1787/a8898906-en.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), 78-94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Gülleroğlu, H.D. (2017). PISA 2012 matematik uygulamasına katılan türk öğrencilerin duyuşsal özeliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesi. *GEFAD / GUJGEF*, *37*(1), 151-175.

Güngör, M., & Atalay Kabasakal, K. (2020). Investigation of measurement invariance of science motivation and self-efficacy model: PISA 2015 Turkey sample. *International Journal of Assessment Tools in Education*, *7*(2), 207-222. https://doi.org/10.21449/ijate.730481

Hangen E. J., & Elliott A.J. (2016) Achievement motives. In: Zeigler-Hill V., Shackelford T. (eds), *Encyclopedia of personality and individual differences.* (pp. 1-3). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_487-1

Helmreich, R. L., & Spence, J. T. (1978). The work and family orientation questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. *JSAS Catalog of Selected Documents in Psychology*, 8(35), Manuscript No. 1677.

Helmreich, R. L., Spence, J. T., Beane, W. E., Lucker, G. W., & Matthews, K. A. (1980). Making it in academic psychology: Demographic and personality correlates of attainment. *Journal of Personality and Social Psychology*, *39*(5), 896–908. https://doi.org/10.1037/0022-3514.39.5.896

Hooper, D., Coughlan, J. & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53-60. https://doi.org/10.21427/D7CF7R

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. http://dx.doi.org/10.1080/10705519909540118

İmrol, F. (2017). *PISA 2012 Türkiye örnekleminde matematiğe yönelik motivasyon ve öz-inanç yapılarının ölçme değişmezliğinin incelenmesi* [Yüksek lisans tezi]. Ankara Üniversitesi.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409-426. https://doi.org/10.1007/BF02291366

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage

Karaman, M. A., & Smith, R. (2019). Turkish adaptation of achievement motivation measure. *International Journal of Progressive Education*, *15*(5), 185-197. https://doi.org/ 10.29329/ijpe.2019.212.13

Karasar, N. (2019). *Bilimsel araştırma yöntemi: Kavramlar ilkeler teknikler.* Nobel Akademik Yayıncılık.

Kaura, N., & Sharma, R. (2015). The effect of gender on achievement motivation. *Indian Journal of Health & Wellbeing*, *6*(5), 504-507.

Khademi, A. (2020). *An investigation of fit criteria within MG-CFA for examining non-negligible measurement invariance* [Unpublished doctoral dissertation, University of Massachusetts Amherst]. Doctoral Dissertations. https://scholarworks.umass.edu/dissertations_2/2035

Khan, Z., Haider, Z., & Ahmed, N. (2011). Gender difference in achievement motivation of intervarsity level badminton players. *Journal of Physical Education and Sport*, *11*(3), 255.

Khojasteh, J. (2012). *Investigating the sensitivity of goodness-of-fit indices to detect measurement invariance in the bifactor model* [Unpublished doctoral dissertation, University of Arkansas]. *Theses and Dissertations*. https://scholarworks.uark.edu/etd/610

Khorramdel, L., Pokropek, A., & van Rijn, P. (2020). Special Topic: establishing comparability and measurement invariance in large-scale assessments, part I. *Psychological Test and Assessment Modeling*, *62*(1), 3-10.

Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye- Çin (Şangay)-Endonezya örneği.* [Yüksek lisans tezi]. Hacettepe Üniversitesi.

Kıbrıslıoğlu Uysal, N, & Akın Arıkan, Ç. (2018). Measurement invariance of science self-efficacy scale in PISA. *International Journal of Assessment Tools in Education*, *5* (2), 325-338. https://doi.org/10.21449/ijate.379508

Kline, R. B. (2015). *Principles and practices of structural equation modeling (4th Ed.).* Taylor & Francis.

Koğar, H., & Koğar, E. Y. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Journal of Measurement and Evaluation in Education and Psychology*, *6*(2). https://doi.org/10.21031/epod.94857

Mascolo, M., & Fischer, K. (1995). Developmental transformations in appraisals for pride, shame and guilt. In J. P. Tangney & K. Fischer (Eds.) S*elf-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 64- 113). Guilford.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*(2), 247–255. https://doi.org/10.1037/0033-2909.107.2.247

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2006, August). *The Utility of Alternative Fit Indices in Tests of Measurement Invariance.* Paper presented at the annual Academy of Management conference, Atlanta, GA.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543. https://doi.org/10.1007/BF02294825

MEB (2019). PISA 2018 Türkiye ön raporu. Retrieved January 2024 from, https://pisa.meb.gov.tr/eski%20dosyalar/wp-content/uploads/2020/01/PISA_2018_Turkiye_On_Raporu.pdf

Nakagawa, S. (2015). Missing data: mechanisms, methods and messages. *Ecological statistics: Contemporary theory and application*, 81-105. https://doi.org/10.1093/acprof:oso/9780199672547.003.0005

Newsom, J. T. (2020). Psy 523/623 structural equation modelling, handouts. Some clarifications and recommendations on fit indices. https://web.pdx.edu/~newsomj/semclass/ho_fit.pdf

Nien, C. L., & Duda, J. L. (2008). Antecedents and consequences of approach and avoidance achievement goals: A test of gender invariance. *Psychology of Sport and Exercise*, *9*(3), 352-372. https://doi.org/10.1016/j.psychsport.2007.05.002

OECD (2019a). *PISA 2018 assessment and analytical framework*. OECD Publishing, Paris. https://doi.org/10.1787/b25efab8-en.

OECD (2019b). *PISA 2018 results (Volume II): Where all students can succeed.* OECD Publishing, Paris. https://doi.org/10.1787/b5fd1b8f-en.

Rosseel, Y., et al. (2022). Package 'lavaan'. Documentation available at CRAN: https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. ASA 1988 Proceedings of the Business and Economic Statistics, Section (308-313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), Latent variables analysis: Applications for developmental research (pp. 399-419). Thousand Oaks, CA: Sage

Schafer, J. L. (1999). Multiple imputation: a primer. *Stat Methods in Med*, *8*(1), 3–15. https://doi.org/10.1191/096228099671525676.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210-222. https://doi.org/10.1016/j.hrmr.2008.03.003

Sehee H., Malik, M.L., & Lee, M. (2003). Testing configural, metric, scalar, and latent mea invariance across genders in sociotropy and autonomy using a non-western sample. *Educational and Psychological Measurement*, *63*(4), 636-654. https://doi.org/10.1177/0013164403251332

Severiens, S., & ten Dam, G. (1998). A multilevel meta-analysis of gender differences in learning orientations. *British Journal of Educational Psychology*, *68*, 595–608. https://doi.org/10.1111/j.2044-8279.1998.tb01315.x

Shekhar, C., & Devi, R. (2012). Achievement motivation across gender and different academic majors. *Journal of Educational and Developmental Psychology*, *2*(2), 105. https://doi.org/ 10.5539/jedp.v2n2p105

Steenkamp, Jan-Benedict E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-90. https://doi.org/10.1086/209528

Steiger, J. H. (1989). *EzPATH: Causal modeling.* SYSTAT. https://www.statpower.net/Steiger%20Biblio/EzPath%20Manual.pdf

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, *43*(4), 599–616. https://doi.org/10.1007/s11135-007-9143-x

Strijbos, J. W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the feedback perceptions questionnaire. *Studies in Educational Evaluation*, *68*, 1-13. https://doi.org/10.1016/j.stueduc.2021.100980

Suna, H. E., Tanberkan, H., & Özer, M. (2020). Changes in literacy of students in Turkey by years and school types: Performance of students in PISA applications. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(1), 76-97. https://doi.org/10.21031/epod.702191

Şekercioğlu, G. (2018). Measurement invariance: Concept and implementation. *International Online Journal of Education and Teaching (IOJET)*, *5*(3). 609-634.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed). Pearson.

Tang, H., & Lu, N. (2013). Measure invariance research of the short form of achievement motive scale in boy-girl of middle school students in Shenzhen. *Advances in Psychology*, *3*(6), 321-326. http://dx.doi.org/10.12677/ap.2013.36048

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). "Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and north Americans": Erratum. *Social Indicators Research*, *78*(2), 341-360. https://doi.org/10.1007/s11205-005-1037-5

Uyar, Ş., & Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi, 2014*(3), 30-43.

Uyar, Ş., & Kaya Uyanık, G. (2019). Fen bilimlerine yönelik öğrenme modelinin ölçme değişmezliğinin incelenmesi: PISA 2015 örneği. *Kastamonu Eğitim Dergisi*, *27*(2), 497-507. https://doi.org/10.24106/kefdergi.2570

van der Sluis, S., Vinkhuyzen, A. A., Boomsma, D. I., & Posthuma, D. (2010). Sex differences in adults' motivation to achieve. *Intelligence*, *38*(4), 433-446. https://doi.org/10.1016/j.intell.2010.04.004

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-69. https://doi.org/10.1177/109442810031002

Wang, C. (2008). *In Search of Diamond Rules* [Unpublished Doctoral dissertation, The Chinese University of Hong Kong]. https://core.ac.uk/download/pdf/48547904.pdf

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26. https://doi.org/10.7275/mhqa-cd89

Yandı, A., Köse, İ. A., & Uysal, Ö. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 13*(1), 243-253. http://dx.doi.org/10.17860/mersinefd.305952

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling A Multidisciplinary Journal 17*(3), 392 – 423. https://doi.org/10.1080/10705511.2010.489003

Yeung, A. S., Craven, R., & Kaur, G. (2012). Gender differences in achievement motivation: grade and cultural considerations. *Psychology of gender differences*, 59-79.