Kastamonu Education Journal, 2025, Vol. 33, No:4, 816-828
doi: 10.24106/kefdergi.1795829

| Research Article / Araştırma Makalesi |

Investigation of Cross-Cultural Measurement Invariance through Multi-Group Confirmatory Factor Analysis and Multi-Group Latent Class Analysis

Kültürlerarası Ölçme Değişmezliğinin Çok Gruplu Doğrulayıcı Faktör Analizi ve Çok Gruplu Örtük Sınıf Analiziyle İncelenmesi[1]

Fatma COŞKUN[2], Hamide Deniz GÜLLEROĞLU[3]

Keywords

1. Measurement invariance
2. Confirmatory factor analysis
3. Latent class analysis
4. TIMSS 2019
5. Student surveys

Anahtar Kelimeler

1. Ölçme değişmezliği
2. Doğrulayıcı faktör analizi
3. Örtük sınıf analizi
4. TIMSS 2019
5. Öğrenci anketleri

Received/Başvuru Tarihi
04.01.2025

Accepted / Kabul Tarihi
10.10.2025

Abstract

Purpose: This study aims to examine cross-cultural measurement invariance using multi-group confirmatory factor analysis (MG-CFA) and multi-group latent class analysis (MG-LCA).

Design/Methodology/Approach: For this purpose, data obtained from the 'Students Like Learning Mathematics Scale' in the TIMSS 2019 study were used. The sample of the research was determined using the maximum variation sampling method. Measurement invariance analyses were conducted on 15 comparison groups formed by Singapore, Hong Kong, Japan, Norway, Turkey, and South Africa.

Findings: The MG-CFA results showed that strict measurement invariance was achieved only between Singapore-Hong Kong and Hong Kong-Norway. Between South Africa and Turkey, measurement invariance was observed at the level of structural invariance, while in other groups, it was achieved at the level of metric invariance. According to the MG-LCA results, measurement invariance was established at the partially homogeneous model level for some groups and at the heterogeneous model level for others.

Highlights: In MG-LCA, it was determined that for the cross-cultural comparisons where measurement invariance was achieved at the partially homogeneous model level, measurement invariance was largely achieved at the metric invariance level according to MG-CFA. The MG-CFA and MG-LCA approaches address measurement invariance from different perspectives. MG-CFA is a sample- and data-oriented approach, whereas MG-LCA is a person-centered approach. Using these two methods together provides more comprehensive insights into why measurement invariance could not be achieved.

Öz

Çalışmanın amacı: Bu çalışmada, kültürlerarası ölçme değişmezliğinin çok gruplu doğrulayıcı faktör analizi (ÇG-DFA) ve çok gruplu örtük sınıf analizi (ÇG-ÖSA) ile incelenmesi amaçlanmıştır.

Materyal ve Yöntem: Bu amaç doğrultusunda, TIMSS 2019 uygulamasında "Matematik Öğrenmeyi Sevme Ölçeği" ile elde edilen veriler kullanılmıştır. Araştırmanın örneklemi, maksimum çeşitlilik yöntemi ile belirlenmiştir. Ölçme değişmezliği analizleri Singapur, Hong Kong, Japonya, Norveç, Türkiye ve Güney Afrika Cumhuriyeti'nin oluşturduğu 15 karşılaştırma grubu üzerinde gerçekleştirilmiştir.

Bulgular: ÇG-DFA sonuçları, katı ölçme değişmezliğinin yalnızca Singapur-Hong Kong ve Hong Kong-Norveç arasında sağlandığını göstermiştir. Güney Afrika Cumhuriyeti ve Türkiye arasında ölçme değişmezliği yapısal değişmezlik düzeyinde kalırken; diğer gruplarda arasında ölçme değişmezliğinin metrik değişmezlik düzeyinde sağlandığı görülmüştür. ÇG-ÖSA sonuçlarına göre, bazı gruplar arasında kısmi homojen model, diğerleri arasında ise heterojen model düzeyinde ölçme değişmezliği sağlanmıştır.

Önemli Vurgular: ÇG-ÖSA' da ölçme değişmezliğinin kısmi homojen model düzeyinde sağlandığı kültürlerarasında ÇG-DFA'ya göre ölçme değişmezliğinin büyük oranda metrik değişmezlik düzeyinde sağlandığı belirlenmiştir. ÇG-DFA ve ÇG-ÖSA yaklaşımları ölçme değişmezliğini farklı boyutlardan ele almaktadır. ÇG-DFA, örneklem ve veri odaklı bir yaklaşımken; ÇG-ÖSA birey merkezli bir yaklaşımdır. Bu iki yöntemin birlikte kullanılması, ölçme değişmezliğinin neden sağlanamadığı hakkında daha kapsamlı bilgiler sağlamaktadır.

[1] This study is based on the doctoral dissertation of the first author under the supervision of the second author.

[2] Corresponded Author, Kahramanmaraş Sütçü İmam University, Faculty of Education, Department of Educational Measurement and Evaluation, Kahramanmaraş, Türkiye; https://orcid.org/0000-0002-6388-3504
[3] Ankara University, Ankara Faculty of Education, Department of Educational Measurement and Evaluation, Ankara, Türkiye; https://orcid.org/0000-0001-7919-0596

Citation/Alıntı: Coşkun, F., & Gülleroğlu, H. D. (2025). Investigation of Cross-Cultural Measurement Invariance through Multi-Group Confirmatory Factor Analysis and Multi-Group Latent Class Analysis, *Kastamonu Education Journal, 33*(4), 816-828. doi: 10.24106/kefdergi.1795829

INTRODUCTION

In recent years, international large-scale assessments have gained increasing importance in educational research. Until the late 1950s, systematic data on educational outcomes were seldom collected at either the national or international level. However, the past decade has seen significant growth in this field (Kirsch & Braun, 2020). Notably, the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), conducted by the International Association for the Evaluation of Educational Achievement (IEA), as well as the Programme for International Student Assessment (PISA), carried out by the Organisation for Economic Co-operation and Development (OECD), are at the forefront of large-scale assessment activities (Hernandez-Torrano & Courtney, 2021). These large-scale assessments capture the attention of the media and the public and play a pivotal role in shaping education policies worldwide (Johansson, 2016). Additionally, many governments make significant financial investments to realign their education policies based on the results of these large-scale assessments (Hambleton & Zenisky, 2011).

Large-scale assessments typically require cross-border measurements conducted through multiple testing tools (Erikcan et al., 2023). In general, large-scale assessments contribute to the development of cross-cultural understanding by comparing student characteristics across different countries (Davidov et al., 2014). One of the primary objectives of large-scale assessments, however, is to compare student characteristics at an international level. This inherently necessitates comparisons across multiple cultures (Hernandez-Torrano & Courtney, 2021). Such comparisons can only be achieved by adapting the tests developed within the scope of large-scale assessments to different languages and cultures (Hambleton & Zenisky, 2011). Tests adapted to various languages and cultures as part of large-scale assessments bring challenges in terms of cross-cultural validity and comparability, such as construct validity, test equivalence, and equivalence of testing conditions (Erikcan et al., 2023). Even when translators achieve absolute accuracy, the cultural context of words used in the source and target languages may lead to differing interpretations (Hambleton & Zenisky, 2011). The lack of equivalence between test versions adapted for different cultures may stem from a variety of factors, including geographic features, traditions, societal moral values, levels of economic development, religious beliefs, or education systems (Hambleton, 2004). To ensure comparable measurement results across different cultures, cultural biases must be minimized (Erikcan et al., 2023). This is achievable through the establishment of measurement invariance. Ignoring measurement invariance significantly jeopardizes the validity of comparison-based analyses (Vandenberg & Lance, 2000).

Failure to properly test the measurement invariance of data obtained from different cultures or nations can lead to the misinterpretation of measurement results (Meitinger et al., 2020). Therefore, before making significant inferences based on cross-cultural data, it is essential to determine whether the relevant constructs are measured in the same way across different cultures (Cieciuch et al., 2019). Van de Vijver (1998) compared comparisons without establishing measurement invariance to comparing apples and oranges, arguing that such comparisons are meaningless and incorrect. Direct comparisons made using terms that may vary across countries, such as the concept of "middle class" commonly used in educational sciences, can result in erroneous conclusions. For instance, the standards of the "middle class" in a country with a high socio-economic status may differ significantly from those in a country with a lower socio-economic status. As a result, studies that do not provide evidence of measurement invariance carry a risk of misinterpreting analysis results (Van de Vijver & Leung, 2021). In this context, education researchers must carefully examine measurement invariance when conducting studies based on international comparisons. For example, Wendt et al. (2017), in a study testing four different levels of measurement invariance using data from 37 countries in the TIMSS and PIRLS applications, found that the model allowing for country-specific measurement structures provided the best fit to the data. This finding highlights the necessity of ensuring measurement invariance for the accuracy of comparative analyses in large-scale applications (Borsboom, 2006).

As understood from the previous explanations, measurement invariance fundamentally refers to the ability of a measurement tool to represent the same construct across all subgroups to which it is applied (Davidov et al., 2014). The literature includes numerous definitions and varied terminologies related to measurement invariance. In his comprehensive study, Johnson (1998) identified more than 50 definitions of "measurement invariance." Additionally, it is observed that the term "measurement equivalence" is also conceptually used in place of measurement invariance in the literature (Kankaras, 2010). Despite the abundance of definitions and varying terminologies in the literature, the common focus of measurement invariance studies lies in the feasibility of making comparisons across groups (Vandenberg & Lance, 2000).

Statistically, measurement invariance is tested within a hierarchical structure, progressing from a model where no parameters are constrained to a model where all parameters are constrained, based on measurements obtained from two different groups (Meredith, 1993; Thissen, 2001). Jöreskog (1971) first investigated statistical evidence for measurement invariance using Multi-Group Confirmatory Factor Analysis (MG-CFA). Subsequently, Sörbom's adaptations of the LISREL program for MG-CFA in 1974 and 1978 contributed to the widespread use of the MG-CFA technique and guided its inclusion in many of today's analysis programs (Van De Schoot et al., 2015). Although statistical analyses of measurement invariance appear to have originated with MG-CFA, many new techniques have been developed over time. Today, numerous methods are employed to test measurement invariance, including MG-CFA, Multi-Group Latent Class Analysis (MG-LCA), Multilevel Confirmatory Factor Analysis, Multilevel Factor Mixture Modeling, Bayesian approaches, Alignment Optimization, and Item Response Theory (IRT)-based models (Kim et al., 2017; Lubke & Neale, 2008). However, the literature shows that MG-CFA, MG-LCA, and IRT-based techniques are the most commonly used methods in measurement invariance studies (Davidov et al., 2014). MG-LCA, in particular, is considered an alternative to MG-CFA by many researchers due to its flexible assumptions (Güngör et al., 2013; Kankaras et al., 2011). MG-LCA was introduced approximately 14 years after the initial use of MG-CFA, in Clogg and Goodman's 1985 paper titled "Simultaneous Latent Structure Analysis in Several Groups." Developed to evaluate measurement invariance in data obtained from different groups, MG-LCA shares many similarities with MG-CFA (Clogg & Goodman, 1985).

In MG-LCA, as in MG-CFA, there is a hierarchical structure progressing from a model where no parameters are constrained across groups to a model where all parameters are constrained. In MG-LCA, the model where all parameters are freely estimated among the compared groups is referred to as the heterogeneous model, the model where certain parameters are fixed is called the partially homogeneous model, and the model where all parameters are fixed is known as the homogeneous model (Clogg & Goodman, 1985; Kankaras & Vermunt, 2014). Despite sharing many theoretical and practical similarities, MG-CFA and MG-LCA also differ in several aspects (Kankaras et al., 2011). For instance, although both techniques are based on latent variable modeling, MG-CFA assumes that latent and observed variables are continuous, whereas MG-LCA treats both latent and observed variables as categorical (Collins & Lanza, 2009; Kankaras et al., 2011). Consequently, MG-CFA requires the fulfillment of univariate and multivariate normality assumptions (Brown & Moore, 2012), while these assumptions are not examined in MG-LCA (Kankaras et al., 2011). MG-LCA operates with more flexible assumptions, with its primary assumption being the principle of local independence (Eid et al., 2003).

From another perspective, while MG-CFA represents a variable-centered approach, MG-LCA is categorized among person-centered approaches (Bergman & Magnusson, 1997; Bergman & Wangby, 2014). This distinction arises from the analytical approaches of the two techniques. In MG-CFA, the focus is on evaluating whether factor structures are similar across groups and identifying potential differences between groups. In contrast, MG-LCA assigns each participant to a specific class based on their response patterns, and the analysis is conducted by examining the differences and similarities among these participant-formed classes (Brown & Moore, 2012; Kankaras et al., 2011). In other words, while MG-CFA involves building a model for a latent trait, MG-LCA identifies latent classes based on participants' response patterns.

When examining measurement invariance studies, it is observed that MG-CFA is the most commonly used technique (Yandı et al., 2015). On the other hand, Eid et al. (2003) have presented compelling arguments regarding the advantages of using MG-LCA to investigate measurement invariance in cross-cultural comparisons. Indeed, numerous subsequent studies (Janousch et al., 2022; Kankaras et al., 2018; Zhao & Jin, 2023) have demonstrated the effectiveness of MG-LCA in cross-cultural measurement invariance studies. Building on this context, this study aims to examine cross-cultural measurement invariance using MG-CFA and MG-LCA techniques with data from TIMSS 2019. A review of the literature reveals that studies employing both techniques together in cross-cultural measurement invariance research are quite limited. From this perspective, this study is expected to contribute to the field. Furthermore, conducting cross-cultural measurement invariance research using TIMSS 2019 data is of significant importance for highlighting the relationship between educational research and cross-cultural measurement invariance. In line with this overarching goal, this study seeks to answer the following questions.

1. Does the Students Like Learning Mathematics Scale ensure measurement invariance across Singapore, Hong Kong, Japan, Norway, Turkey, and South Africa according to the MG-CFA method?
2. Does the Students Like Learning Mathematics Scale ensure measurement invariance across Singapore, Hong Kong, Japan, Norway, Turkey, and South Africa according to the MG-LCA method?

METHOD/MATERIALS

Research Design

This study aims to examine cross-cultural measurement invariance using the Students Like Learning Mathematics Scale (SLLMS) from the TIMSS 2019 assessment, employing MG-CFA and MG-LCA techniques. In this respect, it is characterized as a descriptive study since it seeks to examine the current situation without altering it (Fraenkel & Wallen, 2009).

Study Group

In this study, data from the Students Like Learn Mathematics Scale administered to 4th-grade students in the TIMSS 2019 assessment were used to apply MG-CFA and MG-LCA techniques. Accordingly, the population of this study consists of the 58 countries that participated in the 4th-grade level of TIMSS 2019 (Mullis et al., 2020). However, conducting analyses for all 58 countries in a single study was not feasible due to time and scope limitations. Therefore, countries representing high, medium, and low achievement levels, and differing significantly both geographically and culturally, were selected from among the participating nations. To achieve this, the maximum variation sampling method, a subtype of purposive sampling, was employed, and the analyses were conducted using data from Hong Kong, South Africa, Japan, Norway, Singapore, and Turkey (Gliner et al., 2017). Table 1 presents the student proportions from the countries included in the study sample.

Table 1. Student Proportions of the Countries in the Study Sample

Country	Frequency (f)	Percentage (%)
Hong Kong	3386	9
South Africa	11891	34
Japan	4196	12

Country	Frequency (f)	Percentage (%)
Norway	4526	13
Turkey	4599	13
Singapore	6839	19
Total	35437	100

Data Collection Tools

Students Like Learn Mathematics Scale

SLLMS, included in the TIMSS 2019 assessment, consists of nine items designed to measure students' attitudes toward learning mathematics. The scale is structured as a four-point Likert scale, with items rated as "agree a lot," "agree a little," "disagree a little," and "disagree a lot." Two of the items on the scale are negatively worded and require reverse scoring. This scale was first introduced in the TIMSS assessment in 2011 (Yin & Fishbein, 2020). The data for the SLLMS used in the MG-CFA and MG-LCA analyses in this study were accessed through the IEA's website (https://timssandpirls.bc.edu/timss2019/). The confirmatory factor analysis indices for the SLLMS data used in the study are presented in Table 2.

Table 2. Indices Related to the CFA Results of the SLLMS

Country	χ^2	sd	χ^2/sd	RMSEA	CFI	GFI	SRMR	NFI	IFI
Singapore	1509.09	27	55.89	.09	.96	.95	.03	.96	.96
Hong Kong	1337.89	27	49.55	.12	.94	.92	.05	.94	.94
Japan	1322.54	27	48.98	.11	.95	.93	.05	.95	.95
Norway	874.40	27	32.38	.09	.97	.95	.03	.97	.97
Turkey	1101.28	27	40.78	.09	.94	.95	.05	.94	.94
South Africa	2473.93	27	91.62	.09	.91	.95	.06	.91	.91

Table 2 shows that the fit indices obtained from the CFA of the SLLMS (CFI > .90, GFI > .90, NFI > .90, IFI > .90) indicate an acceptable model fit for the six countries in the study sample (Kline, 2016).

Data Analysis

This study aimed to examine cross-cultural measurement invariance using the data obtained from the SLLMS through MG-CFA and MG-LCA techniques. Before proceeding to the main analyses, the data from Hong Kong, South Africa, Japan, Norway, Singapore, and Turkey were compiled into a single SPSS dataset and examined for missing data and outliers. It was determined that the rate of missing data ranged from 0% to 100% for participants and from 1.6% to 4.4% for scale items. Consequently, 273 participants who left all items unanswered and 1,200 participants with more than 30% missing responses were excluded from the dataset. As recommended by Kline (2016) and Tabachnick and Fidell (2019), the expectation-maximization algorithm was used to address missing data. Univariate outliers were detected based on z-scores, whereas multivariate outliers were identified using Mahalanobis distance (Tabachnick & Fidell, 2019). Following these analyses, data from 33,357 participants were included in the analysis. In tests of normality and multicollinearity assumptions, it was found that the normality assumption was met through skewness and kurtosis coefficients as well as graphical examinations (Demir, 2019). Additionally, multicollinearity assessments indicated that VIF, CI, and tolerance values (VIF < 10; CI < 30; Tolerance > 10) were within acceptable limits (Tabachnick & Fidell, 2019). Exploratory factor analysis was used to assess unidimensionality and local independence. It was determined that the SLLMS exhibited a unidimensional structure, thereby satisfying the assumption of local independence (Demars, 2016). IBM SPSS Statistics (Version 20.0) was used for data preparation. AMOS (Version 24.0) was employed for MG-CFA, and Latent GOLD (Version 6.0) was used for MG-LCA.

The MG-CFA and MG-LCA techniques used to determine cross-cultural measurement invariance were applied to all pairwise comparisons among the six countries comprising the study sample. This approach helps identify varying levels of measurement invariance that may exist among subgroups when full measurement invariance cannot be established across all cultures. Furthermore, it provides more detailed information for cross-cultural comparisons (Davidov et al., 2014). Similarly, Wu et al. (2007), in their study examining cross-cultural measurement invariance using TIMSS data, investigated measurement invariance across 21 comparison groups representing all pairwise comparisons among the seven countries they identified.

In MG-CFA, four levels of measurement invariance were tested, progressing from a model with no parameter constraints to a model with all parameters constrained (Wu et al., 2007). To determine the best-fitting measurement invariance model, several fit indices were considered. The first of these is the "$\Delta\chi^2$" value, based on the difference in chi-square values between two models. While the chi-square difference test ($\Delta\chi^2$) is still widely used, it has been found to be sensitive to sample size. Therefore, it is recommended not to rely solely on the $\Delta\chi^2$ index for decisions regarding measurement invariance (Cheung & Rensvold, 2002; Kline, 2016). In subsequent years, researchers developed the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA), which are less affected by sample size, for use in MG-CFA evaluations (Wu et al., 2007). In this study, the $\Delta\chi^2$, ΔCFI, and ΔRMSEA indices were used to determine the level of measurement invariance in model comparisons during MG-CFA analyses. Accordingly, in model evaluations progressing from structural measurement invariance to strict measurement invariance, at least two of the following criteria were required to be met: ΔRMSEA ≤ .05, ΔCFI ≤ -0.01, and the $\Delta\chi^2$ difference not being significant (Cheung & Rensvold, 2002; Wu et al., 2007).

In MG-LCA, to determine measurement invariance, the first step involved identifying the latent class model representing the six countries in the study sample (Magidson et al., 2020). Determining the number of classes in latent class models is based on principles of parsimony and interpretability, similar to exploratory factor analysis, alongside data reduction (Güngör et al., 2013). Although various indices exist for identifying latent class models, researchers primarily use the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). BIC is more suitable for large samples, whereas AIC is preferable for smaller samples. These indices identify the optimal model based on the number of classes analyzed: the model with the lowest AIC and BIC values as the number of classes increases is considered the best fit (Oberski, 2016). Latent class models, like MG-CFA, also use the chi-square index. However, since the chi-square index is sensitive to sample size in MG-LCA, the use of the BIC index is recommended as an alternative (Magidson et al., 2020). In this study, as the sample sizes analyzed were large enough to affect the chi-square values, the BIC index was used to evaluate the MG-LCA models.

FINDINGS

The findings of the study are presented in two sections. First, MG-CFA was conducted for all pairwise comparison groups among the cultures of Hong Kong, South Africa, Japan, Norway, Singapore, and Turkey, which constitute the study sample. Subsequently, MG-LCA was applied to all comparison groups analyzed in MG-CFA.

Findings Related to MG-CFA

Measurement invariance analyses were conducted across 15 groups based on pairwise comparisons of the countries constituting the sample. These groups are as follows: Singapore-Hong Kong, Singapore-Japan, Singapore-Norway, Singapore-Turkey, Singapore-South Africa, Hong Kong-Japan, Hong Kong-Norway, Hong Kong-Turkey, Hong Kong-South Africa, Japan-Norway, Japan-Turkey, Japan-South Africa, Norway-Turkey, Norway-South Africa, and Turkey-South Africa. Measurement invariance analyses in MG-CFA started by testing the structural invariance model, which imposes the fewest constraints on model parameters. Subsequently, for comparison groups that met the requirements of the next higher model, metric invariance, scalar invariance, and strict invariance models were tested in sequence to conclude the analyses. Table 3 presents the results for the structural invariance model, the initial level of measurement invariance tested in MG-CFA, across the 15 comparison groups.

Table 3. Indices Regarding the Structural Invariance Level of MG-CFA

Comparison Groups	N	χ^2	*p*	RMSEA	CFI
Singapore-Hong Kong	9847	1379.54	.00	.05	.98
Singapore-Japan	10741	1344.11	.00	.05	.98
Singapore-Norway	10812	1477.89	.00	.05	.98
Singapore-Turkey	11110	1300.86	.00	.05	.98
Singapore-South Africa	17359	1256.77	.00	.04	.98
Hong Kong-Japan	7332	1042.35	.00	.05	.98
Hong Kong-Norway	7403	1176.11	.00	.05	.98
Hong Kong-Turkey	7701	999.11	.00	.05	.98
Hong Kong-South Africaa	13950	955.05	.00	.04	.98
Japan-Norway	8297	1440.70	.00	.05	.98
Japan-Turkey	8595	963.69	.00	.05	.98
Japan-South Africa	14844	919.61	.00	.03	.98
Norveç-Turkey	8666	1097.45	.00	.05	.98
Norveç-South Africa	14915	1053.39	.00	.04	.98
Turkey-South Africa	15213	876.35	.00	.03	.98

Table 3 presents the χ^2, RMSEA, and CFI indices to determine whether the structural invariance model is achieved in measurement invariance based on MG-CFA. Upon examining the table, it is observed that the change in the χ^2 difference test is significant for all 15 comparison groups ($p < .05$). However, studies have shown that the χ^2 difference test is prone to significance in MG-CFA conducted with sample sizes larger than 6,000 (Meade et al., 2008), as also evidenced in other research (Cheung & Rensvold, 2002). Therefore, for this study, the CFI and RMSEA indices, which are unaffected by sample size, are more effective in evaluating the measurement invariance models in MG-CFA. Analyzing the information in Table 3, it is determined that the structural invariance model is met for all 15 comparison groups in the study based on the CFI and RMSEA indices. This indicates that the scale structure represented by the SLLMS remains consistent across all comparison groups in terms of the number of subdimensions and the distribution of items to subdimensions (Kline, 2016).

When the structural invariance model is achieved among the groups examined in MG-CFA, analyses proceed to the next level, metric invariance (Wu et al., 2007). Since structural invariance was found to be met for all pairwise comparison groups, the metric invariance model was analyzed for all groups. The MG-CFA results for the metric invariance model are presented in Table 4.

Table 4. Indices Regarding the Metric Invariance Level of MG-CFA

Comparison Groups	ΔCFI	ΔRMSEA	$\Delta\chi^2$	*p*
Singapore-Hong Kong	-0.01	0.00	408.38	.00
Singapore-Japan	-0.01	0.01	565.19	.00
Singapore-Norway	0.00	0.00	153.09	.00
Singapore-Turkey	0.00	0.00	236.51	.00
Singapore-South Africa	-0.01	0.01	675.94	.00
Hong Kong-Japan	0.00	0.00	132.04	.00
Hong Kong-Norway	-0.01	0.00	282.01	.00
Hong Kong-Turkey	-0.01	0.01	524.40	.00
Hong Kong-South Africa	0.00	0.00	100.85	.00
Japan-Norway	-0.01	0.01	442.94	.00
Japan-Turkey	-0.01	0.01	556.79	.00
Japan-South Africa	0.00	0.00	109.52	.00
Norway-Turkey	-0.01	0.00	203.75	.00
Norway-South Africa	-0.01	0.01	524.97	.00
Turkey-South Africa	-0.02	0.01	371.91	.00

Upon examining the information in Table 4, it is observed that the $\Delta\chi^2$ difference test indicates significant changes for all comparison groups ($p < .05$). However, due to the large sample sizes used in MG-CFA, the ΔCFI and ΔRMSEA indices provide more reliable results. Analyzing Table 4, it is evident that the metric invariance model is valid for all comparison groups except Turkey-South Africa based on the ΔCFI and ΔRMSEA indices (ΔCFI ≤ 0.01; RMSEA ≤ 0.015). Accordingly, the MG-CFA results indicate that measurement invariance remains at the structural invariance level for the Turkey-South Africa comparison. Indices for the scalar invariance model are provided in Table 5 for the 14 comparison groups where metric invariance was achieved.

Table 5. Indices Regarding the Scalar Invariance Level of MG-CFA

Comparison Groups	ΔCFI	ΔRMSEA	$\Delta\chi^2$	*p*
Singapore-Hong Kong	-0.01	0.00	408.513	.00
Singapore-Japan	-0.03	0.02	2051.57	.00
Singapore-Norway	-0.02	0.02	1661.08	.00
Singapore-Turkey	-0.04	0.02	2399.19	.00
Singapore-South Africa	-0.05	0.02	3413.36	.00
Hong Kong-Japan	-0.02	0.01	717.13	.00
Hong Kong-Norway	-0.01	0.01	614.16	.00
Hong Kong-Turkey	-0.04	0.02	1434.78	.00
Hong Kong-South Africa	-0.04	0.02	1743.71	.00
Japan-Norway	-0.03	0.02	1585.10	.00
Japan-Turkey	-0.05	0.03	2356.01	.00
Japan-South Africa	-0.08	0.04	4357.66	.00
Norway-Turkey	-0.04	0.02	1996.48	.00
Norway-South Africa	-0.06	0.03	3219.17	.00

Upon examining the information in Table 5, the $\Delta\chi^2$ difference test shows significant changes for all comparison groups ($p < .05$). According to the ΔCFI and ΔRMSEA indices, the scalar invariance model is valid between the cultures of Singapore-Hong Kong and Hong Kong-Norway (ΔCFI ≤ 0.01; ΔRMSEA ≤ 0.015). On the other hand, scalar invariance is achieved for Hong Kong-Japan only based on the ΔRMSEA index. From the information in Table 5, it can be concluded that scalar invariance is satisfied only for Singapore-Hong Kong and Hong Kong-Norway among the 14 comparison groups. Based on this, the next level, strict invariance, was analyzed for these two groups. For the remaining 12 comparison groups, measurement invariance was found to remain at the metric invariance level. The details of the strict invariance model applied to Singapore-Hong Kong and Hong Kong-Norway through MG-CFA are presented in Table 6.

Table 6. Indices Regarding the Scalar Invariance Level of MG-CFA

Comparison Groups	ΔCFI	ΔRMSEA	$\Delta\chi^2$	*p*
Singapore-Hong Kong	-0.01	0.00	444.61	.00
Hong Kong-Norway	-0.01	0.00	415.31	.00

Upon examining the information in Table 6, the Δχ² difference test indicates significant changes between the Singapore-Hong Kong and Hong Kong-Norway comparison groups ($p < .05$). According to the ΔCFI and ΔRMSEA indices, the strict invariance model is valid between the cultures of Singapore-Hong Kong and Hong Kong-Norway (ΔCFI ≤ 0.01; ΔRMSEA ≤ 0.015). Accordingly, it was determined that strict measurement invariance was achieved only between the cultures of Singapore-Hong Kong and Hong Kong-Norway among the 15 comparison groups analyzed using MG-CFA. Structural invariance was found to hold for Turkey-South Africa, while metric invariance was achieved for the remaining 12 comparison groups

Findings Related to MG-LCA

Measurement invariance analyses based on MG-LCA were conducted across 15 groups formed by all pairwise comparisons of the countries in the sample, similar to MG-CFA. To achieve this, it was first necessary to identify the latent class model that best fits the research data for MG-LCA. Accordingly, Latent Class Analysis (LCA) was initially performed to determine the most suitable latent class model for the research data. After identifying the most appropriate latent class model via LCA, the classes within the latent model were named based on their theoretical foundations and the constructs they represent. The process of determining the best-fitting model begins with a single-class latent model and continues by increasing the number of latent classes as long as degrees of freedom allow (Magidson & Vermunt, 2004; Magidson et al., 2020). The model with the lowest information criterion value, used as an evaluation metric, statistically represents the model with the best fit to the data. However, since LCA, like Exploratory Factor Analysis (EFA), is a data reduction method, other criteria such as the number of parameters and the interpretability of the model must also be considered when determining the appropriate model (Arıcıgil-Çilan, 2015; Green, 1952; Güngör-Culha, 2012). Accordingly, LCA was repeated for the six countries in the sample and for the entire dataset to identify the latent class model that best fits the data. The estimated model parameters for models ranging from a single-class model to a six-class model for data obtained through the Students Like Learn Mathematics Scale are presented in Table 7.

Table 7. Model Parameters Related to Latent Class Analysis

Model	Par.[1]	BIC(LL)	AIC(LL)	AIC3(LL)	L^2	sd	Class.[2]
1. Class	27	708239.44	708012.23	708039.23	236485.24	33330	0.00
2. Class	55	588045.86	587583.04	587638.04	116000.04	33302	0.02
3. Class	83	558242.48	557544.04	557627.04	85905.05	33274	0.04
4. Class	111	545867.90	544933.84	545040.84	73238.85	33246	0.06
5. Class	139	539332.54	538162.86	538301.86	66411.86	33218	0.10
6. Class	167	534632.14	533226.83	533393.83	61419.84	33190	0.11

**Note: [1] number of parameters, [2] classification errors*

According to Table 7, as the models progress from a single-class model to a six-class model, there is a noticeable decrease in BIC, AIC, and AIC3 values. This indicates that he data is not limited to a single-class model but instead points to the presence of multiple latent classes. This finding suggests that the data is suitable for latent class analysis (Magidson & Vermunt, 2004). In determining the most appropriate number of classes in the latent model, attention must be given not only to BIC, AIC, and AIC3 information criteria but also to the number of parameters and classification error (Altıntaş, 2016; Lin, 2006). While the decrease in information criteria is pronounced from the single-class model to the three-class model, it becomes much smaller when moving from the three-class model to the six-class model. This pattern can be visualized in a graph with the number of classes on the horizontal axis and information criteria on the vertical axis, similar to the scree plot used in exploratory factor analysis to determine the number of factors (Coşkun, 2023; Coşkun & Gülleroğlu, 2023; Çüm et al., 2020). Considering all this information, the three-class latent model, where the slope of the decrease in information criteria is the lowest, is considered the most reasonable model for explaining the data in terms of the number of parameters and classification error. The three-class latent model, which has a classification error of 4%, indicates that the classification process is largely accurate. Accordingly, the model provides evidence of a 96% likelihood of correct classification (Altıntaş, 2016). The conditional probability parameters of the three-class latent model are presented in Table 8.

Table 8. Conditional Probability Parameters for the Three-Class Latent Model

Item Code	Item Levels	1. Class (0.50)	2. Class (0.33)	3. Class (0.17)
ASBM02A	Strongly Agree	0.83	0.15	0.02
	A little Agree	0.06	0.78	0.16
	A little Disagree	0.02	0.16	0.83
	Strongly Disagree	0.07	0.05	0.88
ASBM02B	Strongly Agree	0.71	0.23	0.06
	A little Agree	0.17	0.58	0.25
	A little Disagree	0.24	0.43	0.33
	Strongly Disagree	0.49	0.17	0.34

ASBM02C	Strongly Agree	0.79	0.17	0.04
	A little Agree	0.19	0.62	0.19
	A little Disagree	0.15	0.51	0.34
	Strongly Disagree	0.40	0.14	0.46
ASBM02D	Strongly Agree	0.75	0.21	0.04
	A little Agree	0.18	0.61	0.21
	A little Disagree	0.09	0.29	0.62
	Strongly Disagree	0.26	0.16	0.58
ASBM02E	Strongly Agree	0.87	0.12	0.01
	A little Agree	0.05	0.87	0.08
	A little Disagree	0.01	0.29	0.70
	Strongly Disagree	0.04	0.04	0.92
ASBM02F	Strongly Agree	0.86	0.12	0.03
	A little Agree	0.33	0.59	0.08
	A little Disagree	0.07	0.49	0.45
	Strongly Disagree	0.12	0.13	0.75
ASBM02G	Strongly Agree	0.87	0.12	0.02
	A little Agree	0.26	0.66	0.08
	A little Disagree	0.05	0.46	0.48
	Strongly Disagree	0.11	0.12	0.77
ASBM02H	Strongly Agree	0.91	0.07	0.02
	A little Agree	0.32	0.64	0.04
	A little Disagree	0.05	0.52	0.43
	Strongly Disagree	0.13	0.11	0.77
ASBM02I	Strongly Agree	0.88	0.11	0.01
	A little Agree	0.28	0.68	0.05
	A little Disagree	0.05	0.63	0.32
	Strongly Disagree	0.06	0.18	0.76

According to Table 8, the class probability values of the three-class latent model are 0.50 for the first class (π_1^x), 0.33 for the second class (π_2^x), and 0.17 for the third class (π_3^x). To describe the characteristics of the three classes identified in the three-class latent model and interpret the behavior of the participant groups, the key parameter to be examined is the conditional probability values calculated based on respondents' answers. Examining Table 8, it is observed that the conditional probability values for the first latent class are concentrated between 0.71 and 0.91 for the "strongly agree" response. Another notable aspect of the first latent class is that the conditional probabilities for "agree a little" and "disagree a little" responses are significantly lower, ranging from 0.02 to 0.33. The answering behavior of respondents in the first latent class resembles response bias, a pattern commonly observed in affective scales (Kankaras, 2010). Another indicator of this behavior is that, despite the presence of two reverse-coded items (ASBM02B, ASBM02B) in the dataset, no change is observed in respondents' answering behavior. A detailed examination reveals that respondents in the first class exhibit an extreme response style (Davidov et al., 2014; Triandis, 1972; Van de Vijver & Leung, 2021). The conditional probability values for the second class in the latent model are concentrated on the "agree a little" and "disagree a little" responses. Based on these conditional probability values, respondents in this class tend to favor moderate options, regardless of the item content. This finding aligns with the "midpoint response style," identified by Van de Vijver and Leung (2021) as a reflection of the humility norm prevalent in East Asia. In the third latent class, respondents predominantly choose the "disagree a little" and "strongly disagree" options. This behavior does not align with any specific response style discussed in the literature. However, examining the conditional probability values for all items suggests that respondents in this class tend to exhibit negative attitudes toward learning mathematics. Based on the characteristics of the three latent classes, they have been named as follows, starting from the first class: "latent class representing extreme response tendency," "latent class representing midpoint response style," and "latent class representing participants with negative attitudes." The distribution of the three latent classes across the six countries examined in the study is presented in Figure 1.

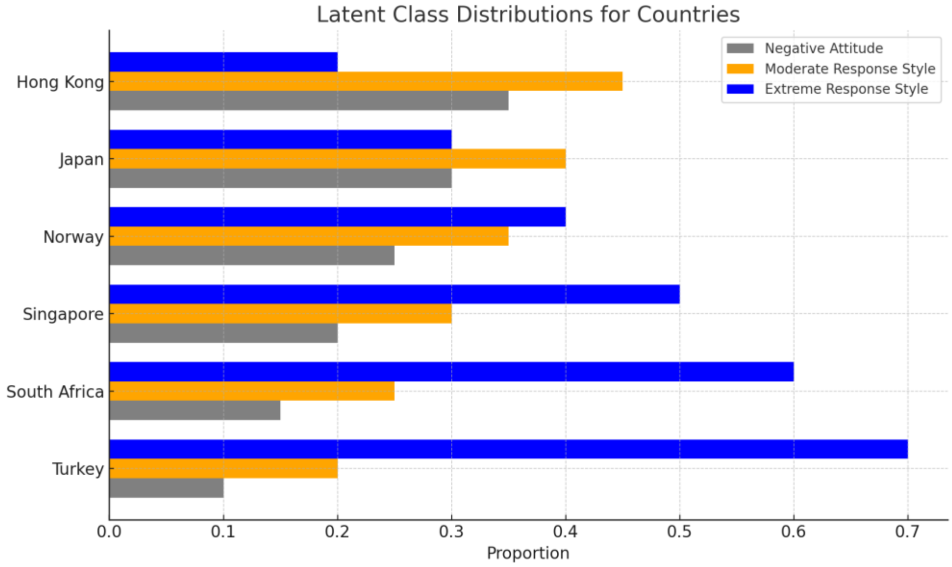

Figure 1. Latent Class Distributions for Countries Included in the Research Sample

According to Figure 1, Turkey and South Africa are predominantly positioned in the first class, representing the "extreme response style," with a probability value of approximately 0.70. Japan stands out in the latent class representing the "midpoint response style," with a probability value of around 0.50. In countries like Hong Kong, Norway, and Singapore, the distribution of respondents across the latent classes appears to be relatively more balanced.

Following the determination and explanation of the most appropriate latent class model for the data examined in the study, both statistically and theoretically, MG-LCA was applied to evaluate measurement invariance. The parameters related to the MG-LCA results for the comparison groups in which measurement invariance was analyzed across all pairwise combinations are presented in Table 9.

Table 9. Model Parameters for MG-LCA Results

Group-Model	BIC(LL)	AIC(LL)	Par.	L^2	df
1-Singapore-Hong Kong					
Heterogeneous Model	167480.84	166286.48	166	35943.88	9681
Partially Homogeneous Model	167381.83	166576.01	112	36341.41	9735
Homogeneous Model	167562.41	166950.84	85	36770.24	9762
2-Singapore-Japan					
Heterogeneous Model	177827.38	176618.59	166	35205.65	10575
Partially Homogeneous Model	177943.93	177128.37	112	35823.42	10629
Homogeneous Model	180742.49	180123.54	85	38872.61	10656
3-Singapore-Norway					
Heterogeneous Model	181128.23	179918.35	166	35453.19	10646
Partially Homogeneous Model	180940.19	180123.89	112	35766.73	10700
Homogeneous Model	182148.43	181528.91	85	37225.75	10727
4-Singpore-Turkey					
Heterogeneous Model	168273.92	167059.53	166	33199.02	10944
Partially Homogeneous Model	168169.95	167350.59	112	33598.09	10998
Homogeneous Model	170720.93	170099.11	85	36400.59	11025
5-Singapore-South Africa					
Heterogeneous Model	283043.48	281755.01	166	55857.77	17193
Partially Homogeneous Model	284540.11	283670.78	112	57881.55	17247
Homogeneous Model	290596.57	289936.81	85	64201.58	17274
6-Hong Kong-Japan					
Heterogeneous Model	123515.51	122370.11	166	28527.63	7166
Partially Homogeneous Model	123356.18	122583.38	112	28848.91	7220
Homogeneous Model	125000.69	124414.19	85	30733.71	7247
7-Hong Kong-Norway					
Heterogeneous Model	126816.86	125669.86	166	28775.16	7237
Partially Homogeneous Model	126699.02	125925.14	112	29138.45	7291
Homogeneous Model	127225.89	126638.57	85	29905.88	7318

Group-Model	BIC(LL)	AIC(LL)	Par.	L^2	df
8-Hong Kong-Turkey					
Heterogeneous Model	113964.58	112811.03	166	26520.98	7535
Partially Homogeneous Model	114011.64	113233.34	112	27051.29	7589
Homogeneous Model	115293.21	114702.54	85	28574.49	7616
9-Hong Kong-South Africa					
Heterogeneous Model	228758.68	227506.51	166	49179.74	13784
Partially Homogeneous Model	229708.49	228863.66	112	50644.89	13838
Homogeneous Model	232611.49	231970.31	85	53805.54	13865
10-Japan-Norway					
Heterogeneous Model	137167.92	136001.99	166	28036.95	8131
Partially Homogeneous Model	137156.49	136369.84	112	28512.81	8185
Homogeneous Model	139695.63	139098.62	85	31295.58	8212
11-Japan-Turkey					
Heterogeneous Model	124314.92	123143.14	166	25782.75	8429
Partially Homogeneous Model	124462.91	123672.31	112	26419.92	8483
Homogeneous Model	128174.05	127574.04	85	30375.65	8510
12-Japan-South Africa					
Heterogeneous Model	239101.09	237838.61	166	48441.49	14678
Partially Homogeneous Model	240398.19	239546.39	112	50257.28	14732
Homogeneous Model	248575.11	247928.65	85	58693.54	14759
13-Norway-Turkey					
Heterogeneous Model	127616.04	126442.89	166	26030.28	8500
Partially Homogeneous Model	127599.07	126807.55	112	26502.94	8554
Homogeneous Model	129845.96	129245.25	85	28994.64	8581
14-Norway-South Africa					
Heterogeneous Model	242401.65	241138.37	166	48689.04	14749
Partially Homogeneous Model	243787.57	242935.24	112	50593.91	14803
Homogeneous Model	249398.86	248752.00	85	56464.67	14830
15-Turkey-South Africa					
Heterogeneous Model	229546.11	228279.54	166	46434.86	15047
Partially Homogeneous Model	230404.02	229549.47	112	47812.79	15101
Homogeneous Model	232199.32	231550.78	85	49868.09	15128

The most commonly used fit indices for determining the level of measurement invariance with MG-LCA are the AIC and BIC information criteria. In MG-LCA evaluations of measurement invariance, the invariance level at which the AIC and BIC information criteria have the lowest value is considered the valid model (Magidson et al., 2020). However, simulation studies have shown that the AIC criterion provides more accurate results for small samples, while the BIC criterion is more effective for large samples (Güngör-Culha, 2012; Kankaras, 2010). Since the data analyzed in this study represent large samples, the BIC information criterion was selected as the appropriate index. Accordingly, examining Table 9, it is determined that partial homogeneity is achieved between Singapore and Hong Kong, Norway, Turkey; Hong Kong and Japan, Norway; and Norway and Turkey, Japan. For the remaining eight comparison groups, measurement invariance was found to remain at the heterogeneous model level.

DISCUSSION, CONCLUSION AND SUGGESTIONS

In this study, cross-cultural measurement invariance was examined using MG-CFA and MG-LCA techniques based on data from the Students Like Learn Mathematics Scale (SLLMS) included in the TIMSS 2019 assessment. Measurement invariance analyses were conducted across 15 comparison groups, encompassing all pairwise comparisons of the cultures of Hong Kong, South Africa, Japan, Norway, Turkey, and Singapore. The MG-CFA results for the 15 comparison groups indicated that strict measurement invariance was achieved only between Singapore-Hong Kong and Hong Kong-Norway. Measurement invariance between Turkey and South Africa was limited to the structural invariance model level, while it remained at the metric invariance level for the other 12 groups. The MG-LCA results revealed that partial homogeneity was established for measurement invariance among the cultures of Singapore-Hong Kong, Singapore-Norway, Singapore-Turkey, Hong Kong-Japan, Hong Kong-Norway, Japan-Norway, and Norway-Turkey. For the remaining 8 groups, measurement invariance remained at the heterogeneous model level. Accordingly, it was found that in MG-LCA, for the cultures where measurement invariance was achieved at the partial homogeneity level—Singapore-Norway, Singapore-Turkey, Singapore-South Africa, Hong Kong-Japan, Japan-Norway, and Norway-Turkey—measurement invariance was achieved at the metric invariance level in MG-CFA. The literature supports that the partial homogeneity model in MG-LCA corresponds to the metric invariance level in MG-CFA (Güngör-Culha, 2012; Kankaras et al., 2011; Yandı et al., 2017). The metric invariance level in MG-CFA indicates that the latent structure under evaluation shows significant similarity across different groups in terms of the number of factors, item factor loadings, and the factors to which the items belong

(Wu et al., 2007). More broadly, the metric invariance model suggests that the regression curves representing the relationship between indicators and the latent variable are equivalent (Kline, 2016). Examining the practical implications of the metric invariance model, Steenkamp and Baumgartner (1998) found that respondents from different groups with metric invariance tend to respond similarly to items or create similar response patterns. In the partial homogeneity model in MG-LCA, it is assumed that the relationship between latent classes and items is consistent across comparison groups. However, the partial homogeneity model does not constrain class distributions and item distributions to remain constant. This implies that latent class structures are similar across comparison groups, but respondent patterns may vary (Kankaras, 2010). In measurement invariance research, the findings obtained from statistical analyses must be examined and explained from both theoretical and practical perspectives. From this viewpoint, the MG-CFA results showing strict measurement invariance between the cultures of Singapore and Hong Kong is an expected outcome. Factors such as the geographical proximity of the two countries, the preference for English as the medium of instruction in both, similarities in educational procedures, and cultural similarities are thought to contribute to achieving strict invariance (Hambleton, 2004; Maden-Kalkan & Yılmaz-Şaşmaz, 2021; Özçelik-Tezel, 2007). An unexpected finding is the establishment of strict measurement invariance between Hong Kong and Norway. Although Hong Kong and Norway are influenced by geographically distinct cultures, such as Northern Europe and Asia, this strict invariance may stem from their shared emphasis on early childhood education, similarities in educational frameworks, and comparable levels of social development (Foy & LaRoche, 2020; Küçükoğlu & Ercan, 2019). The MG-CFA results indicated that Japan achieved metric invariance with all other countries. This suggests that the latent structure examined in the study is similar in Japan and other cultures in terms of the number of factors, the distribution of items across factors, and factor loadings (Kline, 2016). The comparison group with the lowest level of measurement invariance was between Turkey and South Africa, where only structural invariance was achieved. This implies that the latent structure examined in the study shares the same number of factors for Turkey and South Africa, but factor loadings may differ (Kline, 2016).

In addition to MG-LCA, detailed insights into the classes formed by the cultures in the study sample were obtained through latent class analysis (LCA). According to the findings, participants from South Africa, Turkey, and Japan exhibited response patterns similar to certain response styles discussed in the literature. Existing literature emphasizes that response styles substantially impact measurement invariance, particularly in cross-cultural comparisons (Kankaras, 2010; Van de Vijver & Leung, 2021). The MG-LCA results showed that in the comparison groups examined in the study, when both cultures exhibited response styles, there was a 100% likelihood of the heterogeneous model being accepted. Conversely, when neither group exhibited response styles, the partial homogeneity model was accepted with a 100% likelihood. When only one group exhibited a response style, the partial homogeneity model was deemed compatible with a 45% likelihood. From the perspective of MG-CFA, when both countries in the comparison group exhibited response styles, the metric invariance model was identified as appropriate with a 67% likelihood, and the structural invariance model with a 33% likelihood. If only one of the two countries exhibited a response style, the metric invariance model was found to be appropriate with a 100% likelihood. If neither country exhibited response styles, there was a 67% likelihood that the strict invariance model was appropriate.

Overall, the results indicate that when MG-CFA and MG-LCA are used together in cross-cultural studies, they approach measurement invariance from different perspectives and, to some extent, complement each other. It is believed that using these two methods together, rather than relying on a single statistical method, provides more comprehensive insights into questions such as "why measurement invariance could not be achieved." Hambleton (2004) and Borsboom (2006) emphasized that in measurement invariance studies, researchers should not merely determine the statistical level of invariance, as this is an incomplete approach. Instead, the primary objective should be to identify "the reasons why measurement invariance could not be achieved." From this perspective, it was found that response styles identified through latent class analysis approaches also align with the results of MG-CFA. It is suggested that repeating the application of MG-LCA and MG-CFA with different latent traits, samples, and conditions by various researchers would contribute significantly to the field of measurement invariance research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, author-ship, and/or publication of this article.

Statements of publication ethics

We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully.

Researchers' contribution rate

First author conceived of the presented idea and developed the theory. First author performed the computations and conducted the analyses. Second author verified the analytical methods and supervised the research process. Second author also provided critical feedback and guided First author in structuring the study and interpreting the results. Both authors discussed the findings and contributed to the final manuscript.

Ethics Committee Approval Information

The ethics approval for this study was granted by the Ankara University Ethics Committee on August 2, 2021, with approval number 13/276 and reference number 85434274-050.04.04 / 226918.

REFERENCES

Altıntaş, Ö. (2016). *Ankara Üniversitesi yabancı uyruklu öğrenci seçme testinin ölçme değişmezliğinin örtük sınıf ve Rash Modeline göre incelenmesi* (Yayınlanmamış Doktora Tezi), Ankara Üniversitesi, Ankara.

Arıcıgil-Çilan, Ç. (2015). *Uygulamalı gizli sınıf analizi*. İstanbul: Çağlayan Kitabevi.

Bergman, L.R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and psychopathology, 9*(2), 291-319.

Bergman, L.R., & Wangby, M. (2014). The person-oriented approach: A short theoretical and practical guide. *Estonian Journal of Education, 2*(1), 29-49. doi: http://dx.doi.org/10.12697/eha.2014.2.1.02b

Borsboom, D. (2006). When does measurement invariance matter? *Medical care, 44*(11), 176-181.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-379). New York, NY: Guilford Press.

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. doi: http://dx.doi.org/10.1207/S15328007SEM0902_5

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 71*(S1), 157-186. https://doi.org/10.5167/uzh-172539

Clogg, C.C., & Goodman, L.A. (1985). Simultaneous latent structure analysis in several groups. *Sociological Methodology, 15*(1), 81-110.

Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: Wiley. https://doi.org/10.1002/9780470567333

Coşkun, F. (2023). PISA 2018 verisi üzerinden öğrencilerin kültürlerarası iletişim algısının örtük sınıf modelleri ile incelenmesi: türkiye örneklemi. *Gazi Eğitim Bilimleri Dergisi, 9*(1), 1-20.

Coşkun, F., & Gülleroğlu, H. D. (2023). Öğrencileri eğitimde internet kullanımına yönlendirmeye ilişkin öğretmen profillerinin PISA 2018 uygulaması kapsamında karşılaştırması. *Journal of History School, 16*(LXV), 1530-1556. https://doi.org/10.29228/joh.69740

Çüm, S., Demir, E. K., Akın Arıkan, Ç., & Şahin, M. D. (2020). Okuma Zevki Ve Okuma Çeşitliliğinin Örtük Sınıf Analizi: Türkiye Ve Çin İncelemesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 40*(3), 943-977.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., and Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*(1), 55-75. doi: http://dx.doi.org/10.1146/annurev-soc-071913-043137

Demir, E. (2019). *R diliyle istatistik uygulamaları*. Ankara: Pegem Akademi.

DeMars, C.E. (2016). Partially compensatory multidimensional item response theory models: Two alternate model forms. *Educational and Psychological Measurement, 76*(2), 1-27. doi: http://dx.doi.org/10.1177/0013164415589595

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis: A primer. *Journal of Cross-Cultural Psychology, 34*(2), 195-210. https://doi.org/10.1177/0022022102250427

Ercikan, K., Por, H.-H., & Guo, H. (2023). Cross-cultural validity and comparability in assessments of complex constructs. *Educational Testing Service*. https://oecd-ilibrary.org

Foy, P. & LaRoche, S. (2020). Estimating standard errors in the TIMSS 2019 results. In M.O. Martin, M. von Davier, and I.V.S. Mullis, (Eds.) *TIMSS 2019 Technical report*. https://timssandpirls.bc.edu/timss2019/methods

Fraenkel, J.R., & Wallen, N. E. (2009). *How to design and evaluate research in education* (7th ed.). Boston: McGraw Hill Higher Education.

Gliner, J.A., Morgan, G.A., & Leech, N.L. (2017). *Research methods in applied settings: An integrated approach to design and analysis* (Third Edition). New York, London: Routledge Taylor & Francis Group

Green B.F. (1952). Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association, 47*(257), 71-76.

Güngör, D., Korkmaz, M., & Somer, O. (2013). Çoklu-grup örtük sınıf analizi ve ölçme eşdeğerliği. *Türk Psikoloji Dergisi, 28*(72), 48-57.

Güngör-Culha, D. (2012). *Örtük sınıf analizlerinde ölçme eşdeğerliğinin incelenmesi* (Yayınlanmamış doktora tezi). Ege Üniversitesi, İzmir.

Hambleton, R. K., & Zenisky, A. L. (2011). Translating and adapting tests for cross-cultural assessments. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Erlbaum.

Hambleton, R.K. (2004). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. R.K. Hambleton, P.F. Merenda and C.D. Spielberger (Eds). *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-40). London: Psychology Press.

Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: An integrative review and mapping of the literature. *Large-scale Assessments in Education, 9*(17). https://doi.org/10.1186/s40536-021-00109-1

Janousch, C., Sinha, M., & Springer, M. (2022). Resilience profiles across context: A latent profile analysis in a German, Greek, and Swiss sample of adolescents. *PLOS ONE, 17*(1), e0262039. https://doi.org/10.1371/journal.pone.0262039

Johansson, S. (2016). International large-scale assessments: What uses, what consequences? *Educational Research, 58*(2), 139-148. https://doi.org/10.1080/00131881.2016.1165559

Johnson, T. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. In J. A. Harkness (Ed.), Zuma-Nachrichten Spezial Volume 3: *Cross-Cultural Survey Equivalence* (pp. 1–40). Mannheim: Zuma.

Kankaraš M., Moors G., and Vermunt J.K. (2011). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 359-384). New York: Routledge.

Kankaras, M. (2010). *Essays on measurement equivalence in cross-cultural survey research: A latent class approach* (Unpublished PhD thesis). Tilburg University, Netherlands.

Kankaras, M., & Vermunt, J.K. (2014). Simultaneous latent class analysis across groups. A.C. Michalos (Ed.) *Encyclopedia of quality of life and well-being research* (pp. 5969-5974). Heidelberh: Springer.

Kankaraš, M., Moors, G., & Vermunt, J. K. (2018). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis* (2nd ed., pp. 361-384). Routledge. https://doi.org/10.4324/9781315537078

Kankaras, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research, 40*(2), 279-310. https://doi.org/10.1177/0049124111405301

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 524-544. https://doi.org/10.1080/10705511.2017.1304822

Kline, R.B. (2016). *Principles and practice of structural equation modeling* (Fourth Edition). New York, London: The Guilford Press.

Kirsch, I., & Braun, H. (2020). Changing times, changing needs: Enhancing the utility of international large-scale assessments. *Large-scale Assessments in Education, 8*(10). https://doi.org/10.1186/s40536-020-00088-9

Küçükoğlu, M., & Ercan, H. (2019). Norveç'te refah devletinin ortaya çıkışı ve gelişimi. *Uluslararası Toplum Araştırmaları Dergisi, 11*(18), 2275-2308. https://doi.org/10.26466/opus.501680

Lin, H. T. (2006). A comparison of model selection indices for nested latent class models. *Monte Carlo Methods and Applications, 12*(3), 239-259. https://doi.org/10.1515/156939606778705164

Lubke, G., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research, 43*(4), 592-620. https://doi.org/10.1080/00273170802490673

Maden-Kalkan, Ç., & Yılmaz-Şaşmaz, A. (2021). Covid-19 Salgınının Kentsel Yaşam Kalitesi Açısından Potansiyel Etkileri Çin Örneği. *Kent Akademisi, 14*(4), 1283-1298.

Magidson, J., & Vermunt, J.K. (2004). Latent class models. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage Publications.

Magidson, J., Vermunt, J.K., & Madura, J. P. (2020). *Latent class analysis*. London: SAGE Publications Limited.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of applied psychology, 93*(3), 568-592. https://doi.org/10.1037/0021-9010.93.3.568

Meitinger, K., Davidov, E., Schmidt, P., & Braun, M. (2020). Measurement invariance: Testing for it and explaining why it is absent. *Survey Research Methods, 14*(4), 345–349. https://doi.org/10.5167/uzh-192239

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson and M. Kaptein (Eds.), *Modern Statistical Methods for HCI*. Cham, Switzerlans: Springer International Publishing.

Özçelik-Tezel, C. (2007). Ulusal kimliğin oluşumunda müze ve toplum ilişkisi: Singapur. *Elektronik Sosyal Bilimler Dergisi, 6*(20), 133-155.

Steenkamp, J.B.E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research, 25*(1), 78-90.

Tabachnick, B.G., & Fidell, L.S. (2019). *Using multivariate statistics* (Seventh Editon). Boston, MA: Pearson.

Thissen, D. (2001). *IRTLRDIF v. 2.0 b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. Chapel Hill, NC: LL Thurstone Psychometric Laboratory.

Triandis, H. C. (1972). *The analysis of subjective culture*. New York, NY: Wiley.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., and Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology, 6*(1064). doi: https://10.3389/fpsyg.2015.01064

Van de Vijver, F.J., & Leung, K. (2021). *Methods and data analysis for cross-cultural research* (Second Education). United Kingdom: Cambridge University Press.

Van de Vijver, F.J.R. (1998). Towards a theory of bias and equivalence. J.A. Harkness (Ed.), Zuma-Nachrichten Spezial Volume 3. *Cross-Cultural Survey Equivalence* (pp. 41-65). Mannheim: Zuma.

Van de Vijver, F.J.R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. D. Matsumoto and F.J.R. Van de Vijver (Eds.). *Cross-cultural research methods in psychology* (pp. 17–45). New York, NY: Cambridge University Press.

Vandenberg, R.J., and 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), 4-70.

Wendt, H., Kasper, D., & Trendtel, M. (2017). Assuming measurement invariance of background indicators in international comparative educational achievement studies: A challenge for the interpretation of achievement differences. *Large-scale Assessments in Education, 5*(10). https://doi.org/10.1186/s40536-017-0043-9

Wu, A.D., Li, Z., and 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., ve Uysal, Ö. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 13*(1), 243-253.

Yin, L., and Fishbein, B. (2020). Creating and interpreting the TIMSS 2019 context questionnaire scales. In M.O. Martin, M. von Davier, & I.V.S. Mullis (Eds.), Methods and Procedures: *TIMSS 2019 Technical Report* (pp. 16.1-16.331). Boston College, TIMSS & PIRLS International Study Center.

Zhao, M., & Jin, R. (2023). Advancing a cross-cultural understanding of teacher perceptions of school climate: A latent class analysis using 2018 TALIS data. *Frontiers in Psychology, 14*, 1129306. https://doi.org/10.3389/fpsyg.2023.1129306