

An Application of Latent Class Analysis for TIMSS 2015 Data: Detecting Heterogeneous Subgroups

Fatıma Münevver SAATÇIOĞLU *

Abstract

This study aimed to investigate the heterogeneity of the TIMSS 2015 data from Turkey and the USA 8th grade math. Latent Class Analysis (LCA) was used to determine the latent classes that cause heterogeneity in the data by using categorical observed variables. As a result of the LCA, supporting absolute and relative model fit indices through AvePP and entropy values, it was concluded that the data obtained from both countries fit the three-class model. The latent class probabilities and conditional response probabilities were examined for homogeneity and degree of segregation of the classes from each other. Based on the findings, it is recommended that the assumption of homogeneity in international evaluations be evaluated empirically with LCA. With this article, an example of the application of LCA is provided, and it is believed to be useful for researchers in the context of education and psychological evaluation.

Key Words: Latent class analysis, TIMSS 2015, heterogeneity.

INTRODUCTION

The correct understanding of study data is a significant factor for quality research in education and related fields. This is especially true for those investigating the role of scores in latent structures belonging to international large-scale assessment data. In large-scale international assessments such as Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA), item and ability parameters are estimated using Item Response Theory (IRT) calibration. Despite many advantages, IRT models have strict assumptions such as unidimensionality, parameter invariance, local independence, and population homogeneity (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). In order to collect accurate evidence for the validity of the model, the assumptions of the model used in the analysis should be provided and there should be no biased items (Kreiner & Christensen, 2007). In some cases, the population may be heterogeneous due to the techniques or strategies used by individuals to correctly answer the items, familiarity with item content, etc (Embretson, 2007; Mislevy & Huang, 2007). In this case, it is not wise to use statistical models that require a single population in data analysis (Sen, 2016).

Different methods are used for the analysis of data obtained from heterogeneous populations. Analysis methods differ according to whether the population heterogeneity consists of observed or unobserved variables. If the variables causing heterogeneity are observed variables, some of the analysis methods used can be listed as discriminant analysis (DA), logistic regression (LR), multivariate analysis of variance (MANOVA), and multi-group factor analysis (MG-CFA). Of these methods, groups in LR, MANOVA, and MG-CFA are defined using a single observed variable or a combination of observed variables. DA and LR analyses are exploited if the goal is to identify variables to predict group membership, and MANOVA is preferred if it is aimed to compare group means by a set of observed variables. The MG-CFA, on the other hand, is designed for group comparisons by the means and covariances of a set of observed variables. Thus, the MG-CFA includes MANOVA as a submodel. In addition, these methods differ according to the type of observed outcome variables within a

* Lecturer.Dr., Rectorate, Yıldırım Beyazıt University, Ankara-Turkey, fmyigiter@gmail.com, ORCID ID: 0000-0003-4797-207X

To cite this article:

Saatcioglu F.M. (2021). An application of latent class analysis for TIMSS 2015 Data: Detecting Heterogeneous Subgroups. *Journal of Measurement and Evaluation in Education and Psychology*, 12(4), 321-335. doi: 10.21031/epod.984771

Received: 19.08.2021
Accepted: 02.12.2021

subpopulation. If the observed variables are continuous, discriminant analysis (DA) and MANOVA are used, and LR analysis is employed if they are categorical. In MG-CFA, both categorical and continuous observed variables can be included in the same analysis (Lubke & Muthen, 2005).

Analyses using latent variables are person-centered techniques which are K-means clustering analysis, latent class analysis (LCA), and latent profile analysis (LPA) (Lazarsfeld & Henry, 1968; McLachlan & Peel, 2004; Magidson & Vermunt, 2002). Advantage of these person-centered techniques is that they provide a direct analytical translation of theories that hypothesize substantive and qualitative individual differences within the population. These methods are designed to detect clusters of participants with similar response patterns over a set of observed variables in a given dataset. The K-means method is based on an arbitrarily chosen criterion aiming to maximize inter-cluster variability while minimizing intra-cluster variability. LCA and LPA have additional advantages over cluster analysis: (a) individuals are assigned to latent classes based on conditional probabilities, and (b) models are statistically evaluated to decide the most appropriate model based on the observed data (Hagenaars & McCutcheon, 2002). Therefore, LCA and LPA appear in the literature as model-based methods in which alternative models are compared (Vermunt & Magidson, 2002). In LPA and LCA, a single categorical latent variable serves to model class membership (Lazarsfeld & Henry, 1968). For latent variables, analysis methods vary according to the type of observed variable. LPA is used if the latent variable is categorical and the observed variables are continuous, and LCA is favorable if the observed variables are categorical (Lubke & Muthen, 2005). LCA and LPA use multiple observed indicators (i.e., variables) to identify key population subgroups (i.e., latent classes) characterized by different behavioural patterns and are useful when it is not foreknown which participants belong to which subgroups (Butera, Lanza & Coffman, 2014). A latent categorical variable (i.e., underlying class membership) is used to model heterogeneity in the sample (Lubke & Muthén, 2005). In LCA and LPA, all covariation between observed variables is modelled to result from differences between classes. The observed variables within the class are independent of each other, which is called the local independence assumption (Goodman, 2002; McCutcheon, 2002). As it is the only assumption that needs to be met, LCA and LPA are flexible approaches that do not need many assumptions (Lubke & Muthen, 2005; Vermunt & Magidson, 2002). With the LCA, the profiles of the classes are determined by the classes obtained from students with similar reaction patterns (De Ayala & Santiago, 2017). Conditional item probabilities (probability of answering an item for students in a certain class) are used to label latent classes (Nylund, Asparouhov, & Muthén, 2007).

In recent years, latent class modelling techniques have attracted increasing attention among researchers due to the usability and developments in computer software for applications in the social and psychological sciences. Specifically, the use of LCA has increased in many areas such as health (Leech, McNaughton & Timperio, 2014; Olson, Hummer & Harris, 2017) and psychology (Chung, Park, & Lanza, 2005; Collins & Lanza, 2010; Lanza, Flaherty, & Collins, 2003). LCA helps to understand profile differences on multidimensional constructs (like personality, depression, etc.) and provides much more flexibility in parameterizing individual differences. Although LCA is used in various fields such as measurement invariance (Eid, Langeheine & Diener, 2003; Güngör, Korkmaz & Sazak, 2015; Güngör Çulha & Korkmaz, 2011; Kankaras, Moors & Vermunt, 2011; Morin, Meyer, Creusier, & Biétry, 2016; Yandı, Köse & Uysal, 2017), longitudinal latent growth models (Jung & Wickrama, 2008; Rindskopf, 2003) and Differential Item Functioning-DIF (Oliveri, Ercikan, Zumbo, & Lawless, 2014; Samuelsen, 2005; Uyar, 2015).

There are studies investigating latent classes using LCA in large-scale assessments (DeMars & Lau, 2011; Oliveri et al., 2014; Oliveri & von Davier, 2011; Rutkowski, 2018; Toker, 2016), but no exhaustive study has been found on how to apply it step-by-step in TIMSS data. In these studies, latent classes were determined with LCA, and it was indicated that the comments made would cause some adverse conditions due to the fact that IRT assumptions could not be met in the presence of latent classes. First, the presence of more than one latent class in the data obtained from the tests means that the measured structure changes for different classes, and this poses a threat to the validity of the test (Kreiner & Christensen, 2007; Messick 1994; Toker, 2016). It is because providing the assumptions of the model used in data analysis is regarded to be a requirement of ensuring the construct validity (Kreiner &

Christensen, 2007). Second, it is not fair to compare students with the same ability level as the assumptions of the IRT model are violated as a result of detecting latent classes in the data of large-scale assessments (Baghaei & Carstensen, 2013; Embretson, 2007; Oliveri & von Davier, 2011; Rutkowski & Rutkowski, 2018). Another problem is that the parameters estimated using the IRT model may be biased in the presence of different subgroups (DeMars & Lau, 2011; Park, Lee & King, 2016). Therefore, revealing the latent classes that cause heterogeneity in international large-scale test data is highly important in order to be able to analyze the test data correctly and to obtain accurate estimations. In addition, it is hoped that this study will contribute to the literature in terms of providing information on how to apply the LCA analysis to TIMSS 2015 data, how to test its assumptions and how to interpret the analysis outputs.

Purpose of the Study

This study aims to present an example of how to apply the LCA to TIMSS 2015 8th grade math data and to reveal the latent classes.

METHOD

In this study, latent class analysis was used based on students' responses to the items for TIMSS 2015 data. The data were analysed using the maximum likelihood estimation (MLE) method in the Mplus software program (Muthén & Muthén, 2017).

Sample

This study included 432 students from Turkey and 727 students from the USA. In this research, from these countries, the USA was chosen as the country with large sample size while Turkey was chosen as the country with medium sample size. Also, by 8th-grade math achievement average in TIMSS 2015, Turkey ranked 24th, and the USA ranked 10th among 39 OECD countries that took the exam (Mullis, Martin, Foy & Hooper, 2016). Accordingly, it can be alleged that Turkey has a medium level of achievement and the USA has a high level of achievement. So these two countries, which differ in sample size and success ranking, were selected.

Data Collection Tools

TIMSS 2015 is a standardized test that allows 4th and 8th-grade students of countries to determine their knowledge about concepts and processes in math and science, and their attitudes towards these subjects (Thomson, Wernert, O'Grady & Rodrigues, 2017). The instrument of the study is the math achievement test applied to 8th-grade students participating in TIMSS 2015. In the TIMSS assessment, items were developed in accordance with the cognitive processes of knowing, applying, and reasoning. About half of the items in the math test were multiple-choice while the other half consisted of long/short answered items. In TIMSS 2015 with science and math tests, the items in the achievement test included 28 blocks, 14 of which were science and 14 were math. The number of items in the booklets ranged from 11 to 17 (Martin, Mullis & Hooper, 2016). Since the 7th booklet contains more multiple-choice items (17 items), this booklet was chosen and analyzes were carried out. Eight of these items were for measuring knowing, six for applying, and three for cognitive reasoning domains.

Data Analysis

Latent class analysis is one of the finite mixture models used in social, behavioral, and health sciences to determine whether students are divided into latent classes based on a latent structure (Collins & Lanza, 2010). The purpose of LCA is to determine the class membership by using the categorical observed variables. LCA allows the analysis of dichotomously scored (1-0), ordinal and categorical variables, and

the combination of these variables (Nylund, Asparouhov & Muthén, 2007). Figure 1 illustrates below the relationship between latent and observed variables, with the c categorical latent variable (u_1, u_2, u_3) for LCA representing the observed variables:

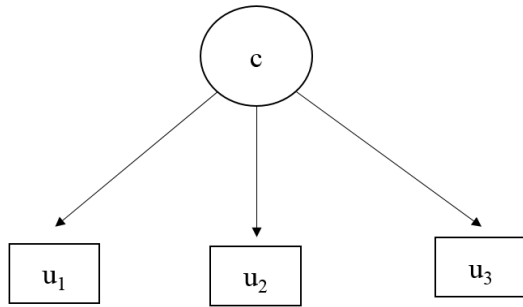


Figure 1. LCA Model for Latent and Observed Variables

The latent variable can be explained as unobservable variables determined by directly measured observed variables. The latent class, on the other hand, represents a statistically determined group of students with homogeneous response patterns, and different latent classes contain different homogeneous response patterns to items (Bolt, Cohen, & Wollack, 2001). In other words, it can be claimed that students in the same latent class have similar abilities, problem-solving skills, and answering strategies (Embretson, 2007; Glück & Spiel, 2007).

If $Y_{ij} \in \{0,1\}$ is the variable showing the responses of individual $i \in \{1,2,\dots,N\}$ to the items; $j \in \{1,2,\dots,T\}$ and $g \in \{1,2,\dots,G\}$ is the variable for the latent class membership of the individual, the probability of answering the items correctly by the individuals in a class $P(Y_{ij} = 1)$ can be equated as follows:

$$P(Y_{ij} = 1) = \sum_{g=1}^G \pi_g P(Y_{ij} = 1|G = g) \quad (1)$$

In this equation, π_g represents the probability of the latent class and the conditional probability of $P(Y_{ij} = 1|G = g)$ demonstrates the probability of answering item j correctly for the individual i in the g class.

As shown in Equation 1, the probability of obtaining an answer of Y_{ij} is the weighted average of the class-specific probabilities. The parameters to be estimated in the latent class model are the latent class probabilities and the conditional response probabilities (Nylund, Asparouhov & Muthén, 2007). These parameters help to examine the degree of homogeneity and latent class separation when evaluating model-data fit. The latent class probability parameters show the population ratio of the students in each latent class. The homogeneity of a latent class means that the students in the class have the same observed response pattern. The fact that the probability of responding to the variables observed in the latent class condition is 0 or 1 gives evidence that the latent classes are homogeneous. The conditional response probability parameters are interpreted while examining the separation of the latent classes. Latent classes are highly differentiated when the conditional response probabilities that are high in one latent class are low in another latent class.

MLE method is used to estimate the latent class analysis parameters. MLE is used to obtain parameter estimates by fitting a given latent class population model to the observed sample data. For mixture models, the likelihood function can generally be obtained by estimating full-information maximum likelihood (FIML) under the assumption of missing-at-random-MAR (McLachlan & Peel 2004). The MLE method continues the estimation of the parameter starting from the initial values until it finds the maximum probability of the parameter. When estimation is not started with appropriate initial values or there is a problem in defining the model, it can give the local maxima value instead of the global maximum of the estimated probability distribution. Estimating the model by taking different random

initial values with the STARTS and STITERATIONS commands added to the syntax in the Mplus software can provide a practical solution to this problem (Jung & Wickrama, 2008; Wang & Wang, 2012). If the problem cannot be dealt with despite taking these precautions, the source of the problem should be determined by examining the defined model, the number of latent classes, observed variables, and sample size.

The assumptions of latent class analysis

Although the latent class analysis does not have assumptions such as normal distribution and unidimensionality, it is necessary to provide the assumption of local independence, indicating that the observed variables are independent. This assumption means that the observed variables are interconnected only with the latent variables and there is no relationship between the errors of the observed variables (Vermunt & Magidson, 2004). To check this assumption, bivariate residuals for observed pairs of variables are examined. High scores of these values suggest local dependency (Vermunt & Magidson, 2004).

Model selection

When the number of classes in a population is unknown in advance, an exploratory approach is followed to determine the number of latent classes. This explanatory approach involves fitting models containing an increasing number of classes to the data and then finding the best fit among these candidate models. As a result of the analysis, the fit indices are compared and the model that best fits the data is selected (Sen, 2016). In determining the number of classes, many factors should be considered, including the research question, parsimony, theoretical justification, and interpretability as well as fit indices (Lubke & Neale, 2006). The principle of parsimony is choosing a model with fewer parameters instead of more complex ones. If latent classes are defined, each latent class must be significant and interpretable. Moreover, even if the model meets all the requirements of mathematical analysis, the predictive model will not be useful if it cannot provide a theoretically interpretable latent class (Wang & Wang, 2012). Therefore, fit indices and model fit tests should not be the decisive factors when deciding on the number of classes. In their simulation study, Nylund et al. (2007) determined that BLRT, followed by BIC, and then sample-size adjusted BIC (SSA-BIC) monitored the best performance among all fit indices and tests found in the Mplus output. However, Nylund et al. (2007) pointed out the disadvantages of the BLRT index due to the increased computation time of BLRT and its dependence on distributional and model assumptions. For example, if the data within a class is skewed but modelled as if the data were normally distributed, the BLRT p -value may be misinterpreted. Hence, Nylund et al. (2007) suggested interpreting the significance of the BIC and SSA-BIC value and the p -value obtained from the VLMR test as a guide to arrive at possible solutions in the first steps of the model research. Therefore, the model with a low BIC and SSA-BIC value and a p -value of less than 0.05 in the VLMR test should be selected (Jung & Wickrama, 2008). The likelihood ratio test (G^2), one of the absolute fit indices, gives information about whether a model fits the data well or not (Agresti, 1990). All of these indices are included in the Mplus output. An additional consideration in model selection is the size of the smallest class. While a four-class model may best fit the data, the researcher should be able to justify the addition of this class if this additional class consists of relatively fewer individuals (e.g. proportionally <1% and/or numerically $n < 25$) (Lubke & Neale, 2006). In addition to all these indices, it is useful to examine the AvePP and entropy values.

The examination of average posterior probabilities and entropy values

When the possible number of classes is optimal, students are assigned to the latent classes in the latent class analysis. Based on a student's response patterns, the probability of latent class membership is measured by the probability of posterior class membership (Wang & Wang, 2012). For this reason, it is very important to examine the mean of posterior probabilities (Average Posterior Probabilities-AvePP) and entropy value related to classification. The posterior mean of probabilities (AvePP) provides

information about how well a particular model classifies students into their classes. Students' AvePP values greater than .70 indicates that the separation of students into classes is successful (Nagin, 2005). The entropy value shows the uncertainty in the classification. A single entropy value is generated for the entire analysis. The entropy value greater than .80 means that the classification uncertainty is low (Clark & Muthen, 2009).

RESULTS

First, it was examined whether the model fit indices and the local independence assumption were provided as a result of the LCA applied to Turkey and the USA data. Then, parameter estimations were examined, and latent class profiles were interpreted based on latent class probabilities and conditional response probabilities.

The Analysis of Model Fit Indices

In the analyses, 1, 2, 3 and 4-class models were tested, respectively, and the obtained model fit indices were presented in Table 1.

Table 1. Fit Indices of Models Tested for Data from Turkey

Fit indices	1-class model	2-class model	3-class model	4-class model
AIC	9004.712	8372.403	8267.709	8253.825
BIC	9073.835	8514.716	8483.212	8542.519
SSA-BIC	9019.887	8403.647	8315.021	8661.097
LR Chi-Square Test	1378.362	58.857	1613.448	193.184
LR Chi-Square <i>p</i> -value	1.0000	1.0000	1.0000	1.0000
VLMR Test	-	668.309	140.694	49.883
VLMR <i>p</i> -value	-	0.0000	0.0000	0.4093
BLRT Test	-	668.309	140.694	49.883
BLRT <i>p</i> -value	-	0.0000	0.0000	0.0200

**p*<.05

According to Table 1, the LR Chi-Square test, which is an absolute fit index, has an insignificant *p*-value for data from Turkey, indicating that the model data fit is achieved. When the BIC and SSA-BIC values of the relative fit indices are examined, it is clear that the 3-class model is the one with the lowest values. The results of VLMR and BLRT can be found in the Mplus output under the Technical 11 and Technical 14 sections, respectively. Here, both VLMR and BLRT show a statistically significant difference between the 2-class and 3-class models. This result implies that the 3-class model provides a significant improvement in model fit compared to the 2-class model. In the next step, the 3-class model is compared with the 4-class model. However, it was observed that the *p* value was not significant when the *p*-value of the VLMR test was examined through testing the 4-class model by adding a class on top of the 3-class model with the suggestion of Nylund et al. (2007). This finding reveals that adding one more class to the 3-class model does not improve the model-data fit. According to these results, it was concluded that the 3-class model fits the data better. The model fit indices of the USA data were submitted in Table 2.

Table 2. Fit Indices of Models Tested for Data from the USA

Fit indices	1-class model	2-class model	3-class model	4-class model
AIC	14653.187	13964.440	13814.329	13788.241
BIC	14731.199	14125.053	14057.542	14114.055
SSA-BIC	14677.219	14013.917	13887.251	13889.608
LR Chi-Square Test	2517.045	2313.400	2204.668	2264.156
LR Chi-Square <i>p</i> -value	1.0000	1.0000	1.0000	1.0000
VLMR Test	-	724.747	186.111	62.088
VLMR <i>p</i> -value	-	0.0000	0.0001	0.3291
BLRT Test	-	724.747	186.111	62.088
BLRT <i>p</i> -value	-	0.0000	0.0000	0.0000

**p*<.05

According to Table 2, the LR Chi-Square test, which is the absolute fit index, has an insignificant *p*-value for the USA data demonstrating that the model data fit is achieved. When the relative fit indices BIC and SSA-BIC are examined, it is observed that the model with the lowest value is the 3-class model. In addition, it was observed that all *p* values were significant at the level of $\alpha = .05$, except for the 4-class model when the *p* values of VLMR tests were examined with the suggestion of Nylund et al. (2007). Therefore, it can be alleged that the model-data fit did not improve as a result of testing the 4-class model by adding a class to the 3-class model. According to the results obtained from the USA data, it was concluded that the 3-class model fits the data better. As a result, latent classes were identified in the TIMSS 2015 8th grade math data from Turkey and the USA, and the heterogeneity in the data was revealed.

The Examination of the Local Independence Assumption

Bivariate residuals (BVR) were examined for observed pairs of variables in testing the local independence assumption, which means that the observed variables are independent in the latent class condition (Vermunt & Magidson, 2004). Higher BVR values (standardized z-score) indicate the presence of local dependency. This information, available in technical output 10 in Mplus software (Muthén & Muthén, 2017), was given in Table 3.

Table 3. Bivariate Residuals (BVR) Examined in 3-Class Model for Turkey and USA Data

Item pairs	TURKEY				USA			
	Category1 Category1	Category1 Category2	Category2 Category1	Category2 Category2	Category1 Category1	Category1 Category2	Category2 Category1	Category2 Category2
M1-M2	-0.355	0.430	0.330	-0.296	0.881	-0.876	-0.669	0.543
M1-M3	0.189	-0.407	-0.370	0.396	-0.451	0.769	0.233	-0.302
M1-M4	-0.187	0.286	0.014	0.008	-0.003	0.139	0.045	-0.104
M1-M5	-0.302	0.477	0.177	-0.210	-0.946	0.743	0.470	-0.596
M1-M6	0.663	-0.907	-0.682	0.546	0.378	-0.380	-0.300	0.242
M1-M7	0.850	-0.759	-0.706	0.687	0.774	-0.512	-0.364	0.255
M1-M8	-0.593	0.929	0.376	-0.476	-0.310	0.562	0.165	-0.223
M1-M9	-0.183	0.204	0.062	-0.041	-0.442	0.413	0.250	-0.267
M1-M10	0.141	-0.275	-0.100	0.123	0.251	-0.103	-0.054	-0.016
M1-M11	0.178	-0.390	-0.204	0.261	0.881	-0.876	-0.669	0.543
M1-M12	-0.027	0.011	-0.003	0.028	-0.451	0.769	0.233	-0.302
M1-M13	-0.377	0.573	0.312	-0.288	-0.003	0.139	0.045	-0.104
M1-M14	0.853	-0.839	-0.775	0.711	-0.946	0.743	0.470	-0.596
M1-M15	0.505	-0.529	-0.516	0.578	0.378	-0.380	-0.300	0.242
M1-M16	-0.451	0.886	0.369	-0.385	0.774	-0.512	-0.364	0.255
M1-M17	-0.472	0.638	0.327	-0.315	-0.310	0.562	0.165	-0.223

When Table 3 is examined, it is observed that the standardized residual values for all variable pairs are near 0. This finding shows that the observed variables in each latent class condition are independent of each other in 3-class model estimation. Accordingly, it is concluded that there will be no local dependency bias in the estimated parameters as there is no local dependency between the observed variables.

The Examination of the Estimated Parameters

The estimated latent class probability parameters and conditional response probability parameters for Turkey data are presented in Table 4. The probability parameters given in parentheses in the table for each latent class represent the population ratio in each latent class. Conditional response probabilities to observed variables under the latent class membership condition were given in Table 4.

Table 4. Parameter Estimations Obtained from 3-Class Model for Data from Turkey

	Levels	Class1 (0.13)	Class2 (0.35)	Class3 (0.52)
M1	0	0.000	0.026	0.588
	1	1.000	0.974	0.412
M2	0	0.130	0.455	0.660
	1	0.870	0.545	0.340
M3	0	0.000	0.514	0.835
	1	1.000	0.486	0.165
M4	0	0.660	0.884	0.787
	1	0.340	0.116	0.213
M5	0	0.471	0.846	0.781
	1	0.529	0.154	0.219
M6	0	0.000	0.111	0.679
	1	1.000	0.889	0.321
M7	0	0.142	0.444	0.620
	1	0.858	0.556	0.380
M8	0	0.419	0.879	0.760
	1	0.581	0.121	0.240
M9	0	0.190	0.506	0.653
	1	0.810	0.494	0.347
M10	0	0.083	0.532	0.784
	1	0.917	0.468	0.216
M11	0	0.356	0.657	0.835
	1	0.644	0.343	0.165
M12	0	0.146	0.812	0.856
	1	0.854	0.188	0.144
M13	0	0.196	0.425	0.771
	1	0.804	0.575	0.229
M14	0	0.006	0.026	0.310
	1	0.994	0.974	0.690
M15	0	0.044	0.097	0.509
	1	0.956	0.903	0.491
M16	0	0.184	0.604	0.854
	1	0.816	0.396	0.146
M17	0	0.246	0.526	0.706
	1	0.754	0.474	0.294

When Table 4 is examined, 13% of the students are in Class 1, 35% are in Class 2, and 52% are in Class 3. According to the conditional response probabilities, the students in Class 1 have a higher performance in answering the items correctly, the students in Class 2 have a moderate performance in answering the items correctly while the students in Class 3 have a lower level in answering the items correctly. The estimated latent class probability parameters and conditional response probability parameters for data from the USA were submitted in Table 5.

Table 5. Parameter Estimations Obtained from 3-Class Model for Data from the USA

	Levels	Class1 (0.18)	Class2 (0.56)	Class3 (0.26)
M1	0	0.021	0.079	0.344
	1	0.979	0.921	0.656
M2	0	0.020	0.319	0.616
	1	0.980	0.681	0.384
M3	0	0.278	0.594	0.719
	1	0.722	0.406	0.281
M4	0	0.091	0.401	0.727
	1	0.909	0.599	0.273
M5	0	0.283	0.842	0.726
	1	0.717	0.158	0.274
M6	0	0.280	0.554	0.744
	1	0.720	0.446	0.256
M7	0	0.069	0.286	0.442
	1	0.931	0.714	0.558
M8	0	0.141	0.450	0.775
	1	0.859	0.550	0.225
M9	0	0.077	0.151	0.403
	1	0.923	0.849	0.597
M10	0	0.140	0.222	0.523
	1	0.860	0.778	0.477
M11	0	0.063	0.134	0.665
	1	0.937	0.866	0.335
M12	0	0.345	0.812	0.876
	1	0.655	0.188	0.124
M13	0	0.010	0.180	0.514
	1	0.990	0.820	0.486
M14	0	0.020	0.203	0.713
	1	0.980	0.797	0.287
M15	0	0.315	0.690	0.698
	1	0.685	0.310	0.302
M16	0	0.652	0.859	0.843
	1	0.348	0.141	0.157
M17	0	0.148	0.561	0.693
	1	0.852	0.439	0.307

When Table 5 is examined, the latent class probability parameters given in parentheses represent the population ratio in each latent class. In other words, 18% of the students are in Class 1, 56% are in Class 2, and 26% are in Class 3. According to the conditional response probabilities, for example, students in latent Class 1 have a higher performance in answering the items correctly, students in Class 2 have a moderate performance in answering the items correctly, while students in Class 3 have a lower performance in answering the items correctly.

The Interpretation of Latent Class Profiles

When the three homogeneous classes obtained from Turkey and the USA data were examined, Class 1, which had a high probability of correctly responding to the items, was interpreted as a *high-performing class*. Class 2, in which the probability of answering the items correctly is moderate, has the characteristics of a *medium-performing class*. Class 3, where the probability of giving correct answers to the relevant items is low, was called the *low-performing class*.

The Examination of Classification Ratios

AvePP and entropy values were calculated to determine the practical usefulness of the model. These values obtained for data from Turkey and the USA were summarized in Table 6.

Table 6. Classification Rates for the 3-Class Model Obtained from Turkey and the USA Data

Turkey		Class 1	Class 2	Class 3
Entropy	0.888	Class 1	0.912	0.088
		Class 2	0.034	0.867
		Class 3	0.000	0.079
The USA		Class 1	Class 2	Class 3
Entropy	0.810	Class 1	0.889	0.110
		Class 2	0.050	0.866
		Class 3	0.000	0.139

When Table 6 is examined, it is observed that the entropy value, which gives an overall value of classification accuracy, is approximately 0.89 for Turkey data and 0.81 for the USA data. It can be asserted that the three-class model is useful in assigning students to the correct classes as the entropy values obtained for both countries are greater than .80 (Clark, 2010). Upon examining the AvePP values, which demonstrate the average of the class probabilities of the students with the maximum posterior probability, it is obvious that these values are above 0.86 for each latent class. These indicate that students have high maximum posterior probability values in being assigned to classes.

DISCUSSION and CONCLUSION

This study presented an example of how to apply the LCA in TIMSS 2015 data, and it was investigated whether the datasets were homogeneous through the LCA. As a result of the LCA, supporting absolute and relative model fit indices through AvePP and entropy values, it was concluded that the data obtained from both countries fit the three-class model. When bivariate residuals (BVR) were examined, it was seen that the observed variables in each latent class condition are independent of each other in 3-class model estimation. The latent class probabilities and conditional response probabilities were reported for homogeneity and degree of segregation of the classes from each other. As a result, the students in Class 1 have a higher performance in answering the items correctly, the students in Class 2 have a moderate performance in answering the items correctly, while the students in Class 3 have a lower level in answering the items correctly. For Turkey, 13% of the students are in Class 1, 35% are in Class 2, and 52% are in Class 3. Also, for the USA 18% of the students are in Class 1, 56% are in Class 2, and 26% are in Class 3. It can be seen that the percentage of the better performing class for American students is more than for Turkish students, while the percentage of the underperforming class for American students is less than for Turkish students. Toker (2016), in his research, examined four countries (Turkey, USA, Finland, Singapore) with different educational systems for TIMSS 2011 8th grade math data. Three latent classes were identified using the latent class analysis. Oliveri et al. (2014) addressed 4th students from Taiwan, Hong Kong, Qatar, and Kuwait who participated in PIRLS 2006. In order to reveal the heterogeneity in the response patterns of the students, three latent classes were determined through the latent class analysis approach. Based on the findings, it is recommended that the assumption of homogeneity in international evaluations be evaluated empirically with LCA.

Also, the indices used during the model determination process in this study, as a result of applying LCA to TIMSS 2015 data, support the simulation results performed by Nylund et al. (2007). It was determined that the BIC and SSABIC values obtained for the number of classes that best fit the model were low. In addition to the model fit indices, the number of classes was decided by examining the p significance value with the VLMR test.

In mixture models such as latent class analysis, the inclusion of auxiliary variables such as covariant variables in the analysis provides valuable information in understanding the population heterogeneity embodied by a latent class variable. In particular, with this approach, it can be determined whether there

are direct effects from covariates to latent variable indicators in an attempt to identify possible sources of DIF (Masyn, 2017).

Future work should focus on extending the classification to include other demographic variables such as gender, age, and socioeconomic status. In addition, it can be used to examine whether latent classes obtained from various distal outcomes (e.g., academic performance, self-efficacy, etc.) show statistically significant mean-level differences or whether these procedures can be included in the latent class determination procedure. Such studies can be used to increase the predictive and discriminant validity of the test. Therefore, they can contribute to test validity.

Latent profile analysis can be used to investigate students' attitude profiles and how these profiles are associated with academic achievement in a standard math and science test for the variables measured by graded Likert-type questionnaires in TIMSS and PISA test data (e.g., attitude). Defined profiles can be a useful way for math and science teachers to understand better the different types of students in their classrooms. Arrangements can be made in the education programs for the deficiencies of the students in the classes.

REFERENCES

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.
- Baghaei, P., & Carstensen, C. H. (2013). Fitting the mixture Rasch model to a reading comprehension test: Identifying reader types. *Practical Assessment, Research, and Evaluation*, *18*(5), 1-13. doi: 10.7275/n191-pt86
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331-348. doi: 10.1111/j.1745-3984.2002.tb01146.x
- Butera, N. M., Lanza, S. T., & Coffman, D. L. (2014). A framework for estimating causal effects in latent class analysis: Is there a causal link between early sex and subsequent profiles of delinquency? *Prevention Science*, *15*(3), 397-407. doi: 10.1007/s11121-013-0417-3
- Chung, H., Park, Y., & Lanza, S. T. (2005). Latent transition analysis with covariates: Pubertal timing and substance use behaviours in adolescent females. *Statistics in Medicine*, *24*, 2895-2910. doi: 10.1002/sim.2148
- Clark, S. L. (2010). *Mixture modeling with behavioral data* (Unpublished doctoral dissertation). University of California, Los Angeles.
- 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. doi: 10.1177/0022022102250427
- Clark, S. L., & Muthén, B. O. (2009). Relating latent class analysis results to variables not included in the analysis. 2009 Manuscript submitted for publication. Retrieved from <http://www.statmodel.com/download/relatinglca.pdf>
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New Jersey: Wiley.
- De Ayala, R. J., & Santiago, S. Y. (2017). An introduction to mixture item response theory models. *Journal of School Psychology*, *60*, 25-40. doi: 10.1016/j.jsp.2016.01.002
- DeMars, C. E., & Lau, A. (2011). Differential item functioning detection with latent classes: How accurately detect who is responding differentially? *Educational and Psychological Measurement*, *71*(4), 597-616. doi:10.1177/0013164411404221
- Embretson, S. E. (2007). *Mixture Rasch models for measurement in cognitive psychology*. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 235-253). New York: Springer Verlag.
- Glück, J., & Spiel, C. (2007). *Studying development via Item Response Model: A wide range of potential uses*. In M. von Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 281-292). New York: Springer Verlag.
- Goodman, L. A. (2002). *Latent class analysis: The empirical study of latent types, latent variables, and latent structures*. In J. A. Hagenars & A. L. McCutcheon (Eds.), *Applied latent class analysis*. New York: Cambridge University Press.
- Güngör Culha, D. & Korkmaz, M. (2011). Örtük sınıf analizi ile bir örnek uygulama. *Eğitimde ve Psikolojide Ölçme Değerlendirme Dergisi*, *2*(2), 191-199.

- Güngör, Culha, D., Korkmaz, M. & Somer, O. (2013). Çoklu-grup örtük sınıf analizi ve ölçme eşdeğerliği. *Türk Psikoloji Dergisi*, 28(72), 48-57.
- Hagenaars, J. A., & McCutcheon, A. (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Jeon, M. (2018). A constrained confirmatory mixture IRT model: Extensions and estimation of the Saltus model using Mplus. *The Quantitative Methods for Psychology*, 14, 120–136. doi: 10.20982/tqmp.14.2.p120
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317. doi: 10.1111/j.1751-9004.2007.00054.x
- Kankaraş, 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, 279–310. doi: 10.1177/0049124111405301
- Kreiner, S., & Christensen, K. B. (2007). *Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models*. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 329-346). New York: Springer Verlag. doi:10.1007/s11336-013-9347-z
- Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). *Latent class and latent transition analysis*. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2, research methods in psychology* (pp. 663-685). Hoboken, NJ: Wiley.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Leech, R. M., McNaughton, S. A., & Timperio, A. (2014). The clustering of diet, physical activity and sedentary behavior in children and adolescents: A review. *International Journal of Behavioral Nutrition and Physical Activity*, 11, 4. doi: 10.1186/1479-5868-11-4
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778. doi: 10.1093/biomet/88.3.767
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10, 21–39. doi: 10.1037/1082-989X.10.1.21
- Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41, 499–532. doi: 10.1207/s15327906mbr4104_4
- Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing*, 20(1), 36–43.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center. Zugriff am (Vol. 21).
- Masyn, K. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling*, 24, 180-197. doi: 10.1080/10705511.2016.1254049
- McLachlan, G. J., & Peel, D. (2004). *Finite mixture models*. New York: Wiley.
- Morin, A. J., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods*, 19(2), 231–254. doi: 10.1177/1094428115621148
- Mullis, I. V., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13–23. doi: 10.3102/0013189X023002013
- Mislevy, R., & Huang, C. W. (2007). *Measurement models as narrative structures*. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications*. New York: Springer Verlag.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th Edition). Los Angeles, CA: Muthén & Muthén.
- Nagin, D. (2005). *Group-based modeling of development*. London: Harvard University.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569. doi: 10.1080/10705510701575396
- Oliveri, M. E., Ercikan, K., Zumbo, B. D., & Lawless, R. (2014). Uncovering substantive patterns in student responses in international large-scale assessments—Comparing a latent class to a manifest DIF approach. *International Journal of Testing*, 14(3), 265-287. doi: 10.1080/15305058.2014.891223
- Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53(3), 315–333.

- Olson J. S., Hummer A. K., & Harris K. M. (2017). Gender and health behavior clustering among U.S. *Young Adults, Biodemography and Social Biology*, 63(1), 3-20. doi: 10.1080/19485565.2016.1262238
- Park, Y. S., Lee, Y.-S., & Xing, K. (2016). Investigating the impact of item parameter drift for item response theory models with mixture distributions. *Frontiers in Psychology*, 7, 255. doi: 10.3389/fpsyg.2016.00255
- Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods*, 8(3), 364–368. doi: 10.1037/1082-989X.8.3.364
- Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research*, 62(3), 354–367. doi: 10.1080/00313831.2016.1261044
- Samuelsen, K. (2005). *Examining differential item functioning from a latent class perspective*. Doctoral Dissertation. Available from ProQuest Dissertations and Theses database. (UMI No. 3175148).
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sen, S. (2016). Applying the Mixture Rasch Model to the Runco Ideational Behavior Scale. *Creativity Research Journal*, 28(4), 426-434. doi: 10.1080/10400419.2016.1229985
- Thomson, S., Wernert, N., O'Grady, E., & Rodrigues, S. (2017). *TIMSS 2015: Reporting Australia's results*. Melbourne, Australia: Australian Council for Educational Research.
- Toker, T. (2016). *A comparison latent class analysis and the mixture Rasch model: A cross-cultural comparison of 8th grade mathematics achievement in the fourth international mathematics and science study (TIMSS-2011)*. Doctoral Dissertation, The Faculty of the Morgridge College of Education University of Denver, USA.
- Uyar, Ş. (2015). *Gözlenen gruplara ve örtük sınıflara göre belirlenen değişen madde fonksiyonunun karşılaştırılması*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Vermunt, J. K., & Magidson, J. (2004). *Latent class analysis*. The Sage Encyclopedia of Social Sciences Research Methods (pp. 549-553).
- Vermunt, J. K., & Magidson, J. (2020). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705511.2020.1818084
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Hoboken, NJ: John Wiley & Sons.
- 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. doi: 10.17860/mersinefd.305952

Appendix. Mplus Code for LCA

```
TITLE:Booklet 7 2 Class Solution Latent Class Analysis
DATA:FILE IS data7.txt;
VARIABLE: NAMES ARE M1-M17;
USEVARIABLES = M1-M17;
CATEGORICAL = M1-M17;
CLASSES = c (2);
MISSING ARE ALL (99);
ANALYSIS:TYPE = MIXTURE;
OUTPUT:TECH1 TECH10 TECH11 TECH14;
SAVEDATA:
FILE IS lca2turkey.dat;
SAVE IS CPROB;
FORMAT IS FREE;
```