

Use of Mixed Item Response Theory in Rating Scales

Seher Yalçın*

Received: 31 October 2018

Revised: 01 November 2018

Accepted: 21 January 2019

ISSN: 1307-9298

Copyright © IEJEE

www.iejee.com

DOI: 10.26822/iejee.2019349251

Abstract

This study aimed to compare the Graded Response Model (GRM) and the Mixed-Graded Response Model (MixGRM) in terms of model data-fit and parameters and demonstrate the application of MixGRM on real data. In this context, this study is basic research based on the International Computer and Information Literacy Study in 2013 conducted with eighth-grade participants from Turkey. The data from a total of 2,356 students were used in the study. In testing the models, data was obtained from an 11-item Likert scale that measured the students' interest and enjoyment in using Information and Communication Technologies (ICTs). When the GRM- and MixGRM-based model data-fit results were compared, the model with the best fit was the MixGRM with four latent classes. Students who reported to enjoy using ICT and who had the highest computer and information literacy (CIL) score were found to be in the first latent class, those with least enjoyment or dislike and those with the lowest CIL score were in the fourth latency class. The findings show that reducing the heterogeneity of Mixed-Item Response Theory models in the dataset is a preferable model for research situations and that Turkish students are not yet prepared for life in the digital age.

Keywords: Graded Response Model, Mixture Item Response Theory, Latent Class Analysis, Mixture Graded Response Model, Information And Communication Technologies

Introduction

Different models and theories are being developed to make decisions about the results of tests taken by individuals more valid and reliable. The theories most often referred to in the literature are; classical test and item response theory (IRT). The classical test theory has certain limitations (Hambleton, Swaminathan & Rogers, 1991), such as being dependent on group, individuals' being dependent on the item they receive, the quality of the item being dependent on the responding group, thus the difficulty of comparing the individuals who take the different tests, being test based and the need for parallel tests for the reliability prediction. Therefore, given these limitations, IRT models are more frequently preferred. Two reasons for this preference are; error prediction is made for each individual to obtain more reliable results, and the item parameters are not changed according to the groups to make the ability prediction independent from the items individuals (De Ayala & Santiago, 2017; De Mars, 2010; Embretson & Reise, 2000). IRT allows predicting the individual's abilities (θ) and item parameters by associating the individual's responses to the item with their ability level with item traits (Embretson & Reise, 2000). Since ability cannot be measured directly, IRT specifies the relationship between the observed success of individuals in the items and the unobservable traits or abilities which are presupposed to lie behind this success (Hambleton & Swaminathan, 1985). However, there are studies in the literature suggesting that IRT-based parameter predictions are more reliable because they create homogeneous latent classes (LCs) according to the response pattern in data in heterogeneous groups, meaning that the sample consists of latent subclasses (De Ayala & Santiago, 2017; Maij-de Meij, Kelderman, & Van der Flier, 2008; Yalçın, 2018). In Mixed-Item Response Theory (MixIRT) models emerging from the combined use of IRT and LC analysis, LCs or homogeneous subgroups are defined. Within each LC, the same measurement model is used, but different parameter estimates between the LCs are undertaken according to IRT (von Davier & Rost, 2017). MixIRT, discovered by Rost in the

1980s, has been widely used in the 2000s. Graded Response Model (GRM) is a logistic model with two parameters for polytomous data (Egberink, Meijer & Veldkamp, 2010). The slope parameter and the threshold parameter representing the number that is one less than the category number are predicted for each item (Embretson & Reise, 2000). The formula for the MixIRT model for GRM is as follows (Egberink et al., 2010; Finch & French, 2012):

$$P(U = 1|g, \theta_{ig}) = \frac{e^{(a_{jg}(\theta_{ig}-b_{jg}))}}{1+e^{(a_{jg}(\theta_{ig}-b_{jg}))}} \quad (\text{Equality 2})$$

where "g: 1, 2, ..., G" represents LC membership, " b_{jg} " indicates the difficulty in the class for the item j, " a_{jg} " shows the in-class selectivity for the item j, and " θ_{ig} " refers to the level of latent ability measured in class for the individual i.

From the studies conducted with MixIRT models in the literature, it can be seen that MixIRT is used in achievement tests (Bolt, Cohen & Wollack, 2001; De Ayala & Santiago, 2017; Yalçın, 2018), Likert-type scales (Egberink et al., 2010; Ölmez & Cohen, 2018), personality questionnaires (Maij-de Meij et al., 2008), and to determine the response style (Eid & Zickar, 2007; Huang, 2016). The literature reveals that more studies were encountered regarding application of MixIRT in the achievement tests but there is a lack of studies on the use and introduction of MixIRT models in graded scales. In one study, MixPCM (Ölmez & Cohen, 2018) was used whereas in another research, Mixed-Graded Response Model (MixGRM) (Egberink et al., 2010) was applied. The current study presents an example of an application through real data and compares the MixGRM and GRM, which are among the MixIRT models for Likert-type items, are frequently used in measuring the latent traits of individuals, such as personality, interest, and attitude. In this context, this paper discusses the students' use of Information and Communication Technologies (ICT), which is considered to have a heterogeneous

*Correspondence Details: Seher Yalçın, Ankara University, Faculty of Education, Department of Measurement and Evaluation, Ankara, Turkey. E-mail: yalcins@ankara.edu.tr

structure and constitute a necessary and indispensable part of their 21st century skills in relation to the literature (Partnership for 21st Century Learning, 2007; Trilling & Fadel, 2009).

The International Computer and Information Literacy Study (Fraillon, Ainley, Schulz, Friedman, & Gebhardt, 2014) reported that the relationship between the highest ICIL and the use of ICT was the strongest in Turkey. Besides, when ranked in terms of the participating countries interest and enjoyment of the use of ICT applications, Turkey is the country with the lowest score in terms of ICIL while it is the third country with the highest score in ICT (Fraillon et al., 2014). In the literature, the attitude toward ICT is related to the student's self-efficiency (Contreras 2004; Güzeller 2011; Rohatgi, Scherer & Hatlevik, 2016; Scherer, Rohatgi & Hatlevik, 2017) toward ICT and computer and information literacy (CIL) (Rohatgi et al., 2016). However, no significant relationship has been determined between the average CIL scores of the students participating in ICILS 2013 from Australia, Germany, Norway and the Czech Republic and having a positive attitude toward ICT, while the relationship was observed to be significant in a negative way in Germany (Gerick, Eickelmann & Bos, 2017). In this context, it seems that there may have been a paradox between some students' attitudes and achievements. In addition, according to the attitude-behavior theory of Fishbein and Ajzen (1975), beliefs regarding a subject lead to an attitude toward it. Beliefs arise from experiences related to the subject and when the beliefs about the subject are positive, the attitude will also be positive. Therefore, beliefs affect our attitude, and our attitude affects our behavior. When considered within the context of the use of information and communication technologies, as students first gain experience with computers, they develop certain beliefs about them (e.g., useful, coercive, entertaining). This makes their attitudes regarding computers to be positive or negative over time and this affects their behaviors toward computers over time; therefore, students can use computers or attempt to avoid using them (Gardner, Dukes and Discenza, 1993). For this reason, it is significant to examine students' attitudes toward ICT use. Moreover, while there are studies (Gerick et al., 2017; Rohatgi et al., 2016) conducted with data from ICILS about different countries, no studies have been encountered in the literature with data on Turkey. For the stated reasons, it is considered necessary to examine students in terms of liking to use ICT by dividing them into homogeneous classes and latent subgroups. This study aims to compare MixGRM with GRM in terms of model data-fit and parameters, and present an example of a MixGRM application on real data by separating students into homogeneous classes in terms of interest and enjoyment regarding the use of ICT. In this context, the following questions will be answered:

1. How does the model data fit the ICT scale for the students that participated in ICILS 2013 from Turkey according to application of MixGRM and GRM?
2. What are the traits of the latent groups according to the model that fits with MixGRM? How are the students distributed in terms of the item response categories for the LCs and the whole group?
3. What are the slope and threshold parameters of the liking ICT use scale according to MixGRM and GRM?

Method

Research Model

This study is basic research since it contributes to the development of theory in terms of the comparison of the parameters predicted according to GRM and MixGRM with

model data-fit results and including application of Mix-GRM on actual data.

Data Source

The research was conducted through the data from ICILS using a two-stage stratified sampling method applied to eighth-grade students participating from Turkey in 2013. The data was obtained from the ICILS webpage (<https://icils.acer.org/>). In this study, from a total of 2540 students that participated in ICILS 2013, analyses were undertaken on the data of a total of 2356 students, 1214 male and 1142 female, after missing data were deleted. The data source in the study consists of students' responses to a Likert-type scale of 11 items that measures students' interest in and enjoyment of ICT within the scope of ICILS 2013.

ICILS 2013, was a comprehensive study implemented out by the International Association for the Evaluation of Educational Achievement (IEA) for the purpose of examining students' ICIL levels. ICILS aims to determine the level of computer and information literacy skills and the factors associated with these skills to support young people's digital age participation capacities in a modern society (Fraillon, Ainley & Schulz, 2013). The participants in ICILS 2013 included 60000 eighth-grade students, 35000 teachers and 3300 schools from 21 education systems of the following 19 countries: Australia, Canada, Chile, Croatia, Czech Republic, Denmark, Germany, Hong Kong, Korea, Lithuania, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Slovenia, Switzerland, Thailand, and Turkey (IEA, 2014).

For the data collection of the study, a national questionnaire was used to obtain information about the education system, and computer and information technology test, student questionnaire, teacher questionnaire, school administrator questionnaire, coordinator questionnaire were employed for international students. In this study, students' gender, computer and information literacy achievement scores and liking to use ICT scale were used. Computer and information literacy achievement scores were converted to a scale with an average of 500 and a standard deviation of 100. Four competence levels for CIL were defined. Level 1 was defined between 407 and 491 points; Level 2 was between 492 and 576 points; Level 3 between 577 and 661 points; and Level 4 for 661 points and above. Students in Level 1 can use traditional software instructions and are familiar with the basic layout rules of electronic documentation. Points below 407 indicate a rather low literacy level even below the targeted Level 1. The average credibility of the liking to use ICT scale for national samples is .81, the factor loadings range from .74 to .86 and having higher points than the scale is interpreted as higher interest and like for ICT (Fraillon, Schulz, Friedman, Ainley & Gebhardt, 2015).

Data Analysis

Firstly, a confirmatory factor analysis (CFA) was conducted in the Mplus-8 program to view the assumption of the one-dimensional of the scale measuring students' interest and enjoyment in using ICT. Information regarding the factor loadings of the items as a result of the analyses are presented in Table 1.

As shown in Table 1, the factor loadings as a result of the CFA were in the range of .678 and .805 and were significant. When the fit indexes obtained from the one-dimensional model were examined, the model appeared to have a good level of fit ($\chi^2_{(44)} = 1329.318, p < .01, RMSEA: .11, CFI: .94, TLI: .93$). For the reliability of the scale, the Cronbach's Alpha Coefficient was calculated as .89. In this context, it

can be stated that the scale is valid and reliable in measuring the interest and enjoyment of students to use ICT.

Table 1. The CFA results of the students' interest and enjoyment in using ICT scale

Items	Estimate	S.E	Est./S.E.
i1	0.698	0.013	54.883*
i2	0.729	0.011	65.126*
i3	0.764	0.012	64.741*
i4	0.805	0.009	92.420*
i5	0.678	0.013	52.627*
i6	0.774	0.010	76.385*
i7	0.757	0.010	75.334*
i8	0.748	0.013	57.640*
i9	0.769	0.010	74.105*
i10	0.743	0.011	67.641*
i11	0.717	0.013	53.762*

For the first sub-goal of the study, a GRM analysis and five MixGRM analyses from the model with one LC to the model with five LCs were practised out in the Mplus-8 program to determine which model fit the data better. For the second question of the study, the index scores regarding the students' gender, computer and information literacy achievement points and the state of liking using ICT were used to determine the characteristics of the latent groups according to the model that fit the liking of using ICT scale. In addition, the distribution of the students according to the response categories were examined and presented in graph format. For the object of finding the answer to the third question of the study, the parameter values were interpreted by presenting according to the model that fits the liking using ICT scale and GRM.

Results

Within the scope of the first purpose of the research, different models were tried in order to determine which model offered the better fit of the data obtained from ICT scale. The results regarding the model fit are presented in Table 2.

Table 2. The results of goodness of fit analyses for the investigated models

Models	AIC	BIC	Entropy
GRM	46250.236	46503.883	-
MixGRM-LC1	46250.249	46503.897	-
MixGRM-LC2	44420.699	44933.759	0.659
MixGRM-LC3	43936.068	44708.541	0.704
MixGRM-LC4	43609.700	44641.585	0.704
MixGRM-LC5	43449.049	44740.346	0.742

As shown in Table 2, when the results of the MixGRM- and GRM-based model data-fit were compared, the model with the best fit for the Bayesian information criterion (BIC) value was MixGRM with four LCs. The entropy value of this model also indicates that the accuracy of the classification was at a good level (Clark, 2010).

Information presented in Table 3 shows the latent group traits according to the best fitting model and reveals the group according to Mix-GRM with four LCs, which was the second sub-goal of the research.

As shown in Table 3, according to MixGRM with four LCs, 48% of the students were in the first LC (LC1), 35% in the second LC (LC2), 9% in the third LC (LC3), and 8% in the

last LC (LC4). When the computer and information literacy score averages of the students were examined, the students with the highest average were in LC1 and the students with the lowest average were in LC4. This was also seen when the indexes were examined in terms of the average of interest and enjoyment regarding the use of ICT. It can be stated that students who were more interested in using ICT were in LC1, and those who less liked or did not like at all were mostly in LC4. While the students in LC1 and LC2 had higher scores in CIL than the average of Turkey (361), the students in the LC3 and LC4 groups had scores that were lower than the average in Turkey. Moreover, when the distribution of the students in the latent classes was examined in terms of sex, it was seen that the rates were almost equal in all the LCs except for LC4, in which 32% were male while 68% were female. According to the findings concerning the latent classes, the students in LC1 had a higher level of CIL and liked to use ICT. LC2 to LC4 students had a gradually decreasing literacy score and level of enjoyment in the use of ICT.

Table 3. Information regarding the Mix-GRM-based model with four LCs

	LC1	LC2	LC3	LC4
Student rate (%)	48	35	9	8
CIL score averages (\bar{x})	375.57	371.90	350.14	307.61
Students' interest and enjoyment to ICT (\bar{x})	57.26	48.92	45.76	42.74
Gender (Female/Male) (frequency)	563/560	405/420	115/108	59/126

The distribution of the response categories of the students according to the LCs and the whole group is presented in Figure 1.

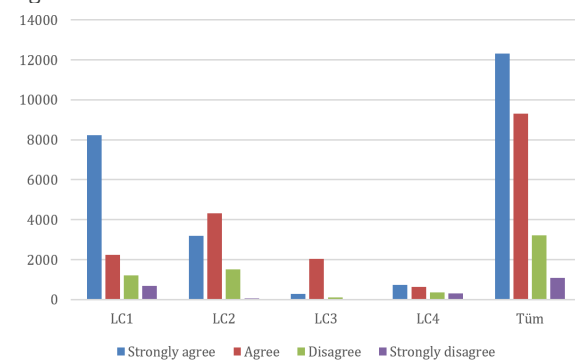


Figure 1. Distribution of response categories according to LCs and the whole group

The students in LC1 had higher responses in the "strongly agree" and "agree" categories than the other LCs (Figure 1). This class can be called the acquiescence response class because the rate of inclusion in the first positive categories was rather higher than the other classes. On the other hand, the "agree" and "disagree" categories in LC2 were in higher ratios than the other classes. Therefore, this class can be called the general class or midpoint in terms of response style. While the category "agree" was dominant in LC3, no dominant category emerged in LC4. When the frequencies in the categories of the students in the whole group were examined, the students usually marked "Strongly agree" and "agree" categories; therefore, this group can be considered as the acquiescence response style.

In the context of the third sub-goal of the study, the results of the analysis conducted to compare the slope and threshold parameters of liking use of ICT scale according to MixGRM and GRM are presented below. First, the slope parameter values predicted according to these models are given in Table 4.

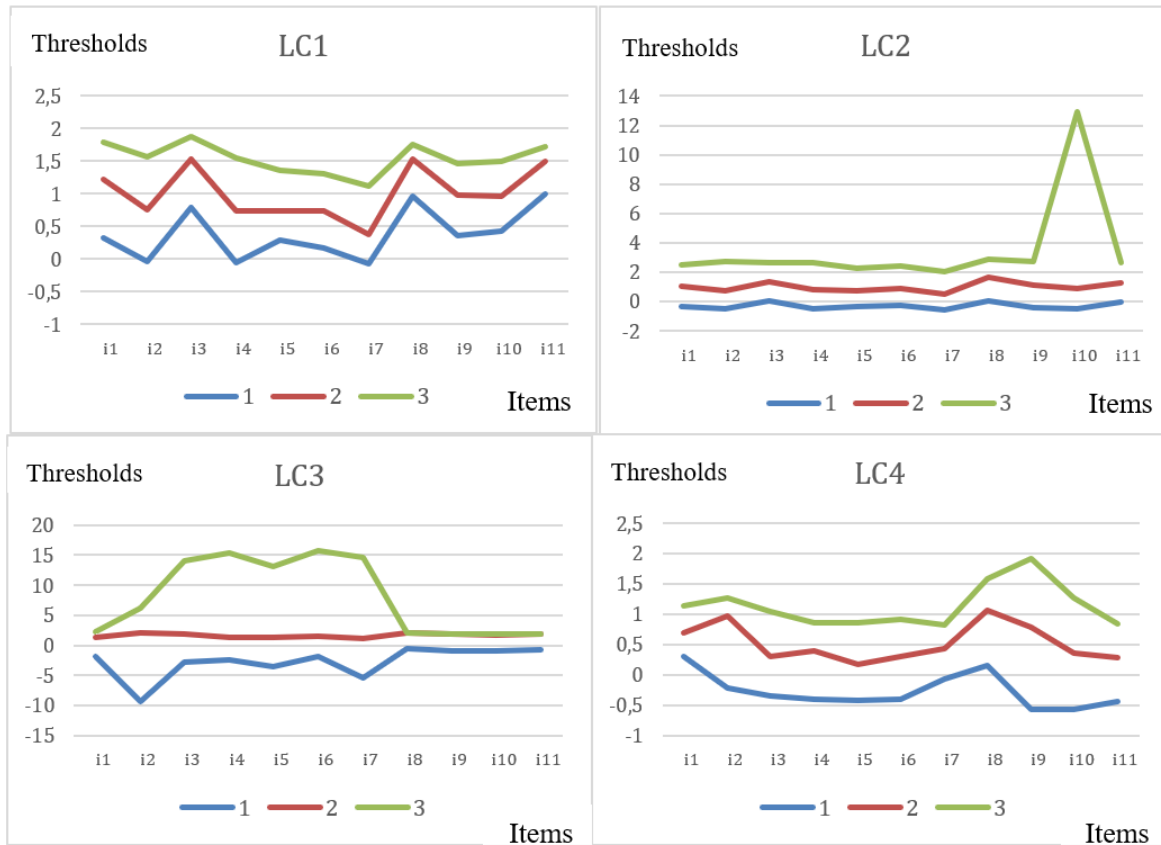


Figure 2. Distribution of the threshold parameter values for LC1, LC2, LC3, LC4

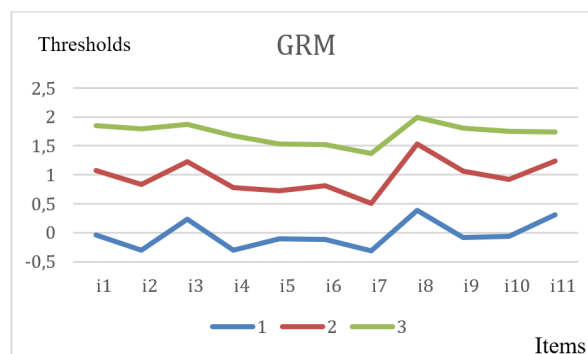


Figure 3. Distribution of the threshold parameter values for GRM

Table 4. The slope parameter values according to different models

Slope parameter (a)	MixGRM				
	LC1	LC2	LC3	LC4	GRM
i1	0.677	0.775	-0.338*	0.702	0.715
i2	0.801	0.713	-0.871	0.721	0.741
i3	0.712	0.804	-0.523*	0.874	0.770
i4	0.849	0.862	-0.233*	0.711	0.816
i5	0.606	0.740	-0.600	0.780	0.684
i6	0.806	0.832	-0.084*	0.717	0.791
i7	0.860	0.828	-0.425	0.553	0.763
i8	0.904	0.675	0.987	-0.002*	0.746
i9	0.887	0.735	0.984	-0.076*	0.770
i10	0.880	0.674	0.993	-0.141*	0.742
i11	0.857	0.665	1.000	-0.027*	0.714
Average	0.804	0.755	0.081	0.437	0.750

*p < .05

When the findings from the item parameters of the LCs in Table 4 were evaluated, it was seen that the highest slope parameter average was in LC1 and the lowest value was in LC3. The average of the slope parameters predicted by GRM was very close to the average of the slope parameters in LC2. The first seven items in LC3 and items 8 to 11 in LC4 had a negative slope parameter value. While items 2, 5 and 7 which had a negative slope value in LC3 were not significant, none of the items with a negative slope value in LC4 was significant. When these items were examined in detail, it was found that the first seven items measured the state of students' finding computers important and entertaining, and their interest in computers, while items 8 to 11 measured students' use of computers, such as doing new things on the computer, using computers for problem solving, and searching for new ways to solve a problem on a computer. The correlations between GRM and the slope parameters predicted for each LC are presented in Table 5.

As shown in Table 5, when the correlations between the slope parameters predicted by GRM and the slope parameters in the LCs were examined, there was only a high level of relationship between LC2 and GRM. The other corre-

lation values were not significant. The distribution of the threshold parameter values predicted according to different models is presented in Figures 2 and Figure 3.

Table 5. Correlation between GRM and slope parameters predicted according to LCs

GRM	LC1	LC2	LC3	LC4
	.314	.651*	.077	.050

As shown in Figures 2 and 3, in all latent classes and GRM, the threshold values gradually increased from “completely agree (1)” to “disagree (4)” for all items. The threshold values indicated where responding to a category was more likely than the previous category. In addition, some threshold values were observed to be negative. This can be interpreted as students’ having a lower level of attitude toward the related items. Concerning the threshold values in the LCs (LC1: 0.978, LC2: 1.41, LC3: 2.36, LC4: 0.47; DTM: 0.884), the lowest average was in LC4 and the highest was in LC3. The correlations between the and the threshold parameters predicted for each LC are presented in Table 6.

Table 6. Correlation between the threshold parameters predicted according to GRM and LCs

GRM	LC1	LC2	LC3	LC4
	.945**	.976**	.900**	.897**

** $p < .05$

When the correlations between GRM-predicted thresholds and LC thresholds were examined, it was found that all the correlations were very high and significant, and that the highest correlation was in LC2. In this context, it can be stated that the estimation by MixGRM had a considerable effect on the slope parameters of the items but not on the threshold parameters.

Discussion, Results and Suggestions

This study aimed to demonstrate the application example of MixGRM on real data by separating students into homogeneous classes with regard to enjoying use of ICT and to compare MixGRM with GRM in terms of model data-fit and parameters. Concerning the goodness of fit results of the GRM- and MixGRM-based models, the model that best fit the data according to the BIC value was MixGRM with four LCs. When the students’ average CIL scores were examined, it was seen that the students with the highest average CIL were in LC1, while those with the lowest average were in LC4. This situation was also observed when the average scores in enjoying use of ICT were analyzed. Besides, the distribution of the students in the LCs according to gender revealed that the results were very similar in all LCs except for LC4. According to the findings, LC1 comprised students that both had higher rates of CIL and enjoyed using ICT. Both the students’ literacy scores and the state of enjoying the use of ICT gradually decreased from LC2 to LC4. In this context, it can be considered that the students in LC4 were those who had the lowest rate of CIL and did not enjoy using ICT or enjoyed it very little. This study demonstrated the relationships between enjoying ICT use and CIL in a linear positive way. In the literature, parallel to this finding, attitudes toward ICT were related to self-efficacy in ICT (Contreras 2004; Güzeller 2011; Rohatgi et al., 2016; Scherer et al., 2017) and CIL (Rohatgi et al., 2016). In addition, self-efficacy in basic ICT skills and CIL achievement were related positively, whereas there was a negative relationship between self-efficacy in advanced ICT skills and CIL (Rohatgi et al., 2016). However, in another study, while the relationship was significant in a negative way only in one country, there was no significant relationship between students’ CIL average and having a positive attitude toward ICT in other countries (Gerick et al., 2017). In this context, it can be stated that there are degrees of

competence perceptions of individuals, and the relations between attitude and achievement could be positive, as well as paradoxical. Therefore, as in this study, it is suggested that the variables that are to be used in related studies, such as attitude, self-efficacy and achievement should be examined by dividing them into sub-categories.

According to the distribution of the students in the response categories, LC1 could be named as the acquiescence response category and LC2 as the general class or midpoint response style. Previous researchers (Eid & Zickar, 2007; Huang, 2016) recommended the use of MixIRT to determine response styles. Furthermore, since different response styles may have different response patterns, MixIRT models can be used to classify individuals according to their response styles (Huang, 2016).

Concerning the item parameters in the LCs, the highest average of the slope parameter was in LC1 and the lowest was in LC3. It was observed that the average slope parameter predicted according to GRM was very close to that found in LC2. When the correlations between the slope parameters predicted according to GRM and the slope parameters in the LCs were examined, it the only significant relationship at a high level was observed between LC2 and GRM. In addition, seven items that measured students’ finding computer important and entertaining, and being interested in computers were negative discriminations for students in LC3, while the items in which students used computers in ways such as doing new things on the computer, using computers for problem solving purposes, and searching for new ways to solve a problem on the computer were classified as negative separators but were not significant for students in LC4. Considering the fact that the students in this LC had the lowest CIL score, this situation can be interpreted as the items measuring such high-level tasks not working at all for this group.

When the threshold averages of LCs were examined, the lowest average was found in LC4 and the highest average in LC3. All correlations between the threshold values predicted according to GRM and the threshold values in LCs were very high and significant, but the highest relation was found with LC2. Furthermore, the threshold values of the items regarding GRM prediction were quite close to the results in the ICILS 2013 technical report (Fraillon et al., 2015). It was concluded that making predictions according to MixGRM significantly affected the slope parameters of the items but did not cause significant differences in the threshold parameters. However, this may have occurred because the threshold value of the items had an influence on the response pattern of individuals. For this reason, it is recommended that this situation is tested under different simulative and actual data conditions.

The obtained findings show that MixIRT models can be selected for studies that require a focus on different subpopulations, such as reducing heterogeneity in the data set and determining the response style, different socio-economic levels, high attitude-low achievement or low attitude-high achievement. Researchers can use MixIRT models in many fields from career development to personality tests. Moreover, when students’ CIL scores are taken into consideration, even those included in the group with the highest score are below Level 1 according to the international score classes, which reveals that Turkish students are not yet prepared for life in the digital age in terms of the 21st century knowledge and media skills.

References

- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response for multiple-choice data. *Journal of Educational and Behavioral Statistics*, 26, 381–409.

- Clark, S. L. (2010). *Mixture modeling with behavioral data* (Unpublished doctoral dissertation). University of California, Los Angeles.
- Contreras, C. L. M. (2004). Predicting computer self-confidence from demographic and personality variables and computer use. *The Quarterly Review of Distance Education*, 5(3), 173-181.
- 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. (2010). *Item response theory*. New York: Oxford University Press.
- Egberink, I. J., Meijer, R. R. & Veldkamp, B. P. (2010). Conscientiousness in the workplace: Applying mixture IRT to investigate scalability and predictive validity. *Journal of Research in Personality*, 44, 232-244.
- Eid, M., & Zickar, M. J. (2007). Detecting response styles and faking in personality and organizational assessments by mixed Rasch models. In M. Von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 255-270). New York: Springer.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Finch, W. H. & French, B. F. (2012). Parameter estimation with mixture item response theory models: A monte carlo comparison of maximum likelihood and bayesian methods. *Journal of Modern Applied Statistical Methods*, 11(1), 167-178. doi: 10.22237/jmasm/1335845580.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Frailon, J., Ainley, J., & Schulz, W. (2013). *International computer and information literacy study: Assessment framework*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Frailon, J., Ainley, J., Schulz, W., Friedman, T. & Gebhardt, E. (2014). *Preparing for life in a digital age: The IEA international computer and information literacy study international report*. Cham: Springer.
- Frailon, J., Schulz, W., Friedman, T., Ainley, J., & Gebhardt, E. (2015). *International computer and information literacy study: ICILS 2013 technical report*. Amsterdam: IEA Secretariat.
- Gardner, D. G., Dukes, R. L., & Discenza, R. (1993). Effects of computer use on computer attitudes and computer confidence: A causal modeling approach. *Computers in Human Behavior*, 9, 427-440.
- Gerick, J., Eickelmann, B., & Bos, W. (2017). The international computer and information literacy study from a European perspective: Introduction to the Special Issue. *European Educational Research Journal*, 16(6), 707-715.
- Güzeller, C. O. (2011). PISA 2009 Türkiye örnekleminde öğrencilerin bilgisayar öz-yeterlik inançları ve bilgisayar tutumları arasındaki ilişkinin incelenmesi. *Kırşehir Eğitim Fakültesi Dergisi*, 12, 183-203.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and application*. Boston: Kluwer Academic Publishers Group.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Huang, H. Y. (2016). Mixture random-effect IRT models for controlling extreme response style on rating scales. *Frontiers in Psychology*, 7(1706), 1-15. doi: 10.3389/fpsyg.2016.01706.
- Maij-de Meij, A. M., Kelderman, H., & Van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, 32, 611-631.
- Ölmez, İ. B. & Cohen, A. S. (2018). A mixture partial credit analysis of math anxiety. *International Journal of Assessment Tools in Education*, 5(4), 611-630. DOI: 10.21449/ijate.455175.
- Partnership for 21st Century Learning (P21). (2007). *Framework for 21st century learning*. Retrieved from <http://www.p21.org/our-work/p21-framework> adresinden alındı.
- Rohatgi, A., Scherer, R. & Hatlevik, O. E. (2016). The role of ICT self-efficacy for students' ICT use and their achievement in a computer and information literacy test. *Computers & Education*, 102, 103-116. Doi: 10.1016/j.compedu.2016.08.001.
- Scherer, R., Rohatgi, A. & Hatlevik, O. E. (2017). Students' profiles of ICT use: Identification, determinants, and relations to achievement in a computer and information literacy test. *Computers in Human Behavior*, 70, 486-499.
- The International Association for the Evaluation of Educational Achievement (IEA). (2014). *Press release*. <https://www.iea.nl/sites/default/files/studies/IEA%20ICILS%202013%20Press%20Release.pdf>
- Trilling, B. & Fadel, C. (2009). *21st century skills: Learning for life in our times*. Francisco: Jossey-Bass.
- von Davier, M. & Rost, J. (2017). Logistic mixture-distribution response models. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume one: Models* (pp. 393-406). Boca Raton: Chapman and Hall/CRC.
- Yalçın, S. (2018). Data fit comparison of mixture item response theory models and traditional models. *International Journal of Assessment Tools in Education*, 5(2), 301-313. doi: 10.21449/ijate.402806.