Attitudes toward Pirated Content: A Scale Development Study Based on Graded Response Model

Eren Can AYBEK[1], H. Deniz GULLEROGLU[2]

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

Purpose: The present study aims to develop a scale based on the Graded Response Model, one of the Item Response Theory models to determine university students' attitudes towards pirated content. Most of the data collection tools on attitudes towards piracy and pirated content have not been developed as a scale, and an attitude scale towards pirated content based on Item Response Theory (IRT) does not exist in the literature.

Research Methods: In accordance with this purpose, the data collected from 438 university students studying at two public universities were used in this study. First, the estimation of the Item Response Theory was tested, and then item parameters were estimated according to the Graded Response Model by analyzing the item-model compatibility.

Findings: The findings obtained in this study showed that a 17-item Likert-type unidimensional scale was developed based on the Graded Response Model to determine the university students' attitudes towards pirated content. Upon the analysis of the information function of the scale, the findings showed that the scale provided more information for individuals between trait levels -1.5 and 1.5 and that the scale provided a very reliable measurement for the individuals with trait levels -2 and 3. Moreover, the scale showed measurement invariance and DIF-free according to gender.

Implications for Research and Practice: Using this scale in further studies on pirated content, group-independent measurement results can be obtained. Furthermore, through studies, such as differential item functioning (DIF) and computerized adaptive testing (CAT), the practicality of the scale can be increased.

[1] Pamukkale University, TURKEY, e-mail: erencan@aybek.net, ORCID: 0000-0003-3040-2337
[2] Ankara University, TURKEY, e-mail: denizgulleroglu@yahoo.com, ORCID: 0000-0001-6995-8223

Introduction

The term piracy is generally used to describe the deliberate infringement of copyright on a commercial scale. Internet piracy or digital piracy is defined as unauthorized use of music or other content on the internet (International Federation of the Phonographic Industry [IFPI], 2013). In other words, digital piracy is to copy or download copyrighted products, such as software, movies, and audio files (Yoon, 2011, p. 405). Gopal et al. (2002) define digital piracy as the illegal copying of any digital product without the permission of the copyright holder (cited in Yu, 2012, p. 364). Software piracy, as a part of digital piracy, is defined as unauthorized copying or distribution of copyrighted software (BSA, 2013), and music piracy, which is another type of digital piracy, is described as unauthorized copying and reproduction of recorded music CDs for commercial purposes (MU-YAP, 2013). Based on all these definitions in the literature, pirated content can be defined as copying and distributing the contents, such as music, video, books, software in a physical or digital environment without the permission of the copyright holder.

When studies on pirated content are analyzed, some studies analyze the university students' justification of digital piracy (Yu, 2012) and how advertising can help stop university students from downloading illegal music (Sheehan et al., 2012). However, there are also studies about the factors affecting the attitude towards digital piracy (Al-Rafee & Cronan, 2006), demographic variables that affect university students' attitudes towards software piracy (Acilar, 2010) and the gender roles in the attitude towards digital piracy (Okurame & Ogunfowora, 2011). Pastore and Cesareo (2014) developed a scale to measure the attitudes and behaviors of consumers towards online music piracy and tried to measure the consumers' attitude through a five-item subscale. The internal consistency coefficient for this subscale was .91. Dilmperi et al. (2017) developed a 28-item scale in a study on online music piracy. Four items of this scale measure attitude towards online music piracy.

Determining the attitudes of individuals towards pirated content will play an important role in the formulation of strategies and policies to be developed in this context. However, what exactly is attitude? Thurstone (1928) defines attitude as the "sum of the individual's tendencies and feelings, prejudices or biases, established thoughts, ideas, emotions, things that threaten the individual, and their views on any subject" (p. 531). Based on the definitions of pirated content and attitude, the attitude towards pirated content can be defined in the context of this research as *an individual's tendencies, prejudices, biases, settled thoughts, ideas and emotions related to pirated music, video, book or software.*

As the title of Thurstone (1928) suggests, *attitudes can be measured*. Measurement process runs through scales, and traditionally the scales are developed based on Classical Test Theory (CTT). However, the psychometric characteristics of the scale items developed according to the CTT are based on the group in which the scales are implemented, and the characteristics of the group depend on the characteristics of the scale items (Hambleton & Swaminathan, 1985, p. 2). For example, an item on a depression scale would have high discrimination for a highly heterogenous group, yet

the same item would have low discrimination if we use it amongst depressed people only, which is a highly homogenous group. Another test development theory is the Item Response Theory (IRT). Since the person and the item characteristics are represented on the same scale, IRT exceeds the limitations of CTT and it makes possible to the prediction of an individual's response to an item to a certain extent and allows us to make individual error estimations (de Ayala, 2009, p. 5). These features of IRT enable the development of computerized adaptive tests (CAT) and provide novel methods and techniques for test equating and bias studies. Such studies need scale items that are calibrated to IRT. Researchers are able to calibrate scales that were developed under the CTT to IRT with a prior study. However, we should note that selecting an IRT calibrated scale for these kinds of studies would be more useful.

The models developed for IRT are categorized according to item scoring, size of the trait measured by the item, or whether it is parametric or not. Applying IRT models may vary whether the item scoring methods are dichotomous and polytomous or the item size is unidimensional and multidimensional. Dichotomous scored item response theory models are used for the items scored as 1-0, whereas polytomous scored item response theory models are applied with the items scored other than 1-0. The models used for the dichotomous items are Rasch model, One Parameter Logistic (1 PL), 2 PL and 3 PL models; some of the models used for the polytomous items are Graded Response Model (GRM), Partial Credit Model (PCM) and Generalized Partial Credit Model (GPCM). In this study, polytomous item response theory models were preferred as a five-point Likert type scale was to be developed.

It has been observed in the studies on pirated content that generally used questionnaires as a data collection tool (Acilar, 2010; Al-Rafee & Cronan, 2006; Okurame & Ogunfowora, 2011). In these studies, no scale was developed; therefore, no validity and reliability studies were conducted. Thus, it is not possible to obtain total scores from these instruments and to perform transactions on total points. The scale used by Pastore and Cesareo (2014) was developed according to the Classical Test Theory (CTT) for online piracy only. Similarly, Dilmperi et al. (2017) developed the attitude scale according to the CTT, which aims to measure attitudes towards online music piracy only.

There are many scale development studies based on IRT in the literature. Rubio, Aguado, Hontangas and Hernández (2007) developed an emotional adjustment measure under the GRM. They indicated that the scale is appropriate for theta levels between -2 and 2. Demirtasli, Yalcin and Ayan (2016) also developed an attitude scale towards educational measurement courses based on the GRM. They developed a 33-items scale, and the scale provides more information for respondents whose theta levels between -2 and 2. Yasar and Aybek (2019) developed a GRM based scale towards resilience of the university students. They developed 95 draft items as a 7-point Likert-type scale, and after the calibrations, the final form of the scale consisted of 35 items. When they examined the item information function, they indicated that the scale gives more information to the respondents with theta level between -2.50 and 2.50. The marginal reliability coefficient has been found as .973. In another study, Yasar (2019) developed a valid and reliable scale that measures perceived stress with GRM.

These are only a small part of the current literature on the scale development studies based on the Graded Response Model.

As an individual's behavior plays an important role in the policies to be adopted by anti-piracy organizations or states, a scale to determine the individuals' attitudes towards pirated content is needed. Previous studies tried to fulfill this need with scales developed under the CTT. However, there was no IRT calibrated scale to measure attitudes towards pirated content. The present study aims to develop a scale to investigate the university students' attitudes towards pirated content, and which the measurement results are reliable and valid based on the item response theory.

Method

Research Sample

The data in this study were collected from 494 university students studying in the faculties of education at two public universities during the 2016-2017 academic year. However, after excluding invalid forms and extreme values from the data set, - 438 university students (89 women, 346 men, and three participants who did not specify gender) were reached (130 sophomores, 55 junior and 253 senior years).

Scale Development Process

The scale was developed based on the IRT model called GRM (Samejima, 1996). In GRM, comparisons are made between groups of response categories and probabilities of a respondent's endorsement, a group of categories calculated as 1 vs. 2-3-4-5; 1-2 vs. 3-4-5; 1-2-3 vs. 4-5, and 1-2-3-4 vs. 5 for five-point Likert response categories (i.e., cumulative comparisons are made in GRM). Before starting the development of the Attitude Scale toward Pirated Content (ASPC), according to GRM, a scale development plan (Baykul, 2010, p. 313) was made and the psychological structure was defined as individual's tendencies, prejudices, biases, settled thoughts, ideas and emotions related to pirated content by the literature review about the structure intended to be measured by this scale. Afterwards, in addition to the literature search for writing the scale items, 60 university students outside the research sample were asked to write essays about pirated content. In the essay, they were asked about their opinions about downloading (or streaming via illegal websites) movies, TV series, and music; and downloading cracked software and e-books from illegal ways, such as private or public torrent trackers. Pirated content was defined for all respondents before the application. According to the responses to the essay, 17 cognitive, 11 affective and 25 behavioral, 53 items were developed and 17 of the items were negative (anti-piracy endorsement) items. Draft items were analyzed by three measurement and evaluation specialists, and the items that do not include any expressions related to attitude, the ones which could not measure the desired or which had ambiguous expressions, were removed, and a 17-item scale draft form was created.

During writing the items and creating the scale draft, it was ensured that the number of items in the positive and negative expressions of the scale was close to each

other. Nine negative (anti-piracy endorsement) and eight positive (piracy endorsement) items were included in the scale. The distribution of the items was as 10 cognitive, three affective, and four behavioral. This distribution seemed far from the distribution of the first draft items (17 cognitive, 11 affective, and 25 behavioral). This was because the specialists' suggestions on most of the affective and behavioral items were phenomenon statements.

Draft form of the scale applied to a group that consisted of 177 university students. This group was not included in the main sample. When the response of the pilot group was analyzed, it was that the scale was unidimensional, and all the items had discriminant indexes, which had higher than .40. Therefore, there were no changes made to the scale items. Since the IRT calibration study required larger samples, the scale was applied to the main sample of this study, which was different from the pilot group and consisted of 438 students. The scale had a form of a five-point Likert-type which included "Strongly Disagree, Disagree, Partly Agree, Agree and Strongly Agree", and it consisted of 17 items.

Data Collection

The draft scale was applied to the university students in paper and pencil form. Demographic variables were also collected during the application to describe the research sample. The students who were involuntary to participate in this study were excluded from the research sample and the application lasted approximately 10 minutes.

Data Analysis

Data preparation

The data collected from the research sample were prepared for analysis, missing values were analyzed, and it was realized that, in a 494 x17 matrix, 39 cells in total consisted of a missing value. Since the ratio of the amount of missing value to the whole data was quite small, the rows that contained the missing value were removed from the data set and the analysis continued with the data that were collected from 457 people. After that, whether there was an extreme value in the data set was determined by identifying the Mahalanobis distances. After calculating Mahalanobis distances, the significance level of the chi-square value for 16 degrees of freedom (number of items -1) was analyzed. The values found as $p < .001$ were accepted as extreme values and excluded from the data set. After excluding the extreme values, 438 people were included in the data set. To check whether there was a multicollinearity problem between the items, the correlation values between the items were calculated. Due to the highest correlation value is $r=.79$ in the analysis, it was decided that there was no multicollinearity between the items.

IRT assumptions

The principal axis factoring technique was chosen for the exploratory factor analysis as this study aims to reveal a latent structure (Field, 2009, p. 638). In the case of more than one factor, it was decided to use the varimax technique which is one of

orthogonal rotation techniques. The reason why the vertical rotation method was not preferred was that the aim was not to adjust the results of the analysis with the existing data but generalize (Alpar, 2011, p. 213). While deciding the number of factors, both relationships between factor eigenvalues and scree plot graph were considered. The analyses were conducted using IBM SPSS Statistics v20 for exploratory factor analysis to check the assumption of IRT. Also, Yen's Q3 statistics examined to test the local independence assumption. mirt package (Chalmers, 2012) is used to obtain the Yen's Q3 statistics.

IRT calibrations

R 3.5.0 (R Core Team, 2013) software with mirt package (Chalmers, 2012) was used for IRT calibrations. Chalmers' (2012) mirt package is intended to handle IRT analysis. It is capable of run dichotomous and polytomous IRT models, such as GRM, GPCM and NRM. GRM is used to IRT calibrations for this study. Also, option response functions (ORFs), item information functions, test information function, and reliability function were plotted with mirt. Since there are several options which respondents could choose, unlike the dichotomous models, there are characteristic curves for each option, and these curves are named as ORF (de Ayala, 2009).

Measurement invariance and differential item functioning

As an extra validity proof, measurement invariance was tested among gender groups using lavaan (Rosseel, 2012) and semTools (Jorgensen et al., 2020) packages in R. As an indicator of the invariance, $\Delta CFI \leq .01$ was used, which is recommended by Cheung and Rensvold (2002) and Chen (2007). Differential Item Functioning (DIF) analysis was also run to investigate whether the items showed the same response function among genders or not. DIF analysis was run by lordif (Choi, 2016) package in R. Because the data structure was polytomous, a logistic ordinal regression method was used. As the DIF detection criteria, the p-value for χ^2 less than .01 and change in pseudo $R^2 \geq .13$ (Zumbo, 1999) was used.

Results

Unidimensionality, Exploratory Factor Analysis Results

Exploratory factor analysis was conducted to investigate whether unidimensionality and local independence assumptions were fulfilled, and items were analyzed to see if they were under a single dominant factor. Before exploratory factor analysis, correlation coefficients between items were analyzed and it was found that the highest correlation coefficient was r = .79 (p <.05). The absence of a correlation coefficient over .90 among the items indicated that there were no replicates or items that measured the same characteristics. Subsequently, item discrimination indexes were calculated, and it was concluded that the lowest item discrimination index was .432. Therefore, all items were included in exploratory factor analysis and Kaiser-Meyer-Olkin (KMO) and Bartlett Sphericity test results were analyzed to investigate whether the data set was suitable to conduct exploratory factor analysis.

The KMO value was .92 and the Bartlett Sphericity test result was significant (χ^2 = 3559.897; p <.05), which showed that the data set was suitable for exploratory factor analysis. In the exploratory factor analysis, the principal axis factoring was used as the factor extraction technique, and the findings showed that according to the scree plot (Figure 1), 17 items were collected under a single factor.
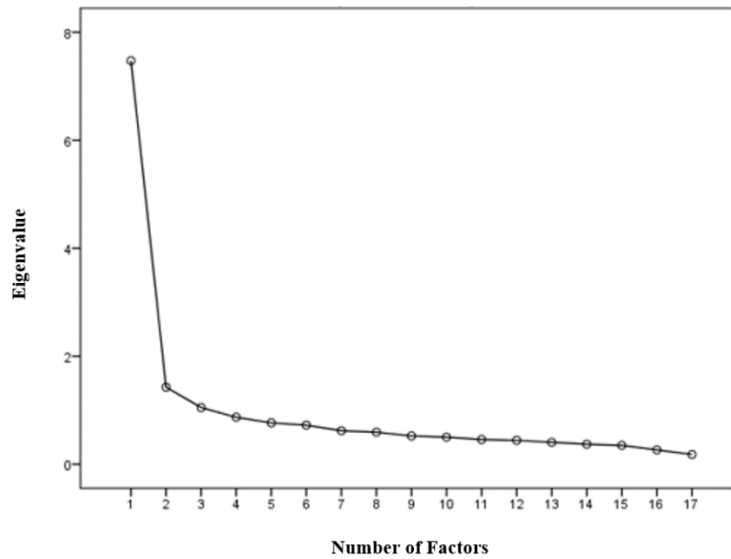

Figure 1. *Scree Plot for Exploratory Factor Analysis of 17 Items*

As the scree plot indicates, a single latent variable could explain the variances of the 17 items. Table 1 shows the factor loadings of the items, the amount of variance by single factor and the contribution of each item to the commonality. According to Table 1, the lowest factor loading was .446 (Item 16) and the lowest contribution to the commonality was .199 (Item 16). In addition, the findings showed that 40.71% of the total variance could be explained by a single factor. This shows that the items provide the assumption of unidimensionality.

To investigate whether local independence was achieved or not, the Q3 statistics proposed by Yen (1993) were calculated. Correlation values between residual values were calculated for each pair of items. Cohen d effect size limit values were used to evaluate Q3 statistics. Accordingly, it can be inferred that the Q3 value between 0.24 and 0.36 was moderate and over 0.37 can be interpreted as a high level of dependence (Smits et al., 2012). Accordingly, it was found that there was a moderate dependence between the items 2-3; 2-7; 11-13; 11-16; 12-14 and 15-16. There were not any items that were highly dependent. Furthermore, as the unidimensionality was ensured, no interventions were made to pairs of moderately dependent items. As a result of these findings, it was accepted that the data provided the assumptions of unidimensionality and local independence.

Table 1

Exploratory Factor Analysis Results of the Attitude Scale toward Pirated Content (ASPC)

Item	Contribution to commonality	Factor Loading	r_{jx}
1	.313	.559	.538
2	.482	.694	.658
3	.428	.654	.619
4	.260	.510	.497
5	.377	.614	.592
6	.452	.673	.638
7	.363	.602	.583
8	.235	.485	.459
9	.566	.752	.716
10	.584	.764	.731
11	.647	.804	.760
12	.395	.629	.603
13	.603	.776	.736
14	.409	.639	.611
15	.394	.628	.607
16	.199	.446	.432
17	.215	.464	.448

KMO = .928; χ^2 = 3559.897; p < .05

Explained variance = %40.704

Cronbach's alpha = .92

r_{jx}: Item Discrimination Index

Model-Data Fit

When item calibrations were performed for the Graded Response Model, the Log-Likelihood value of the model was 9088.368. In addition, the marginal reliability coefficient calculated to determine the reliability of predicted ability levels (de Ayala, 2009, p. 177) was .946. S-χ^2 statistics were used to determine whether the items fit with the model. LaHuis et al. (2009) stated that the S-χ^2 statistic has lower Type I error than expected in their studies, which they compared different item fit statistics.

Accordingly, it was concluded that all items were fit to the model (p> .05). Item fit statistics are given in Table 2.

Findings about Item Calibrations

After it was found that the scale fit with the Graded Response Model, discrimination parameter and category boundary location was calculated for each item and these parameters are given in Table 2.

Table 2

Item Parameters and Item Fit Indices of ASPC Items

Item Number	a	b_1	b_2	b_3	b_4	$S\text{-}\chi^2$	$sd_{S\text{-}\chi^2}$	$p_{S\text{-}\chi^2}$
1	1.504	-1.263	-0.395	-0.795	1.990	90.437	90	.467
2	2.284	-1.105	0.041	0.857	1.612	56.503	73	.923
3	2.056	-1.618	-0.435	0.474	1.424	57.828	73	.903
4	1.301	-1.209	0.174	1.532	2.686	91.836	86	.314
5	1.725	-1.473	-0.246	0.801	2.311	75.072	73	.411
6	2.120	-1.142	0.072	0.772	1.834	68.965	75	.674
7	1.772	-0.265	1.050	1.959	2.789	61.875	56	.274
8	1.239	-1.282	0.256	1.055	2.375	100.202	94	.312
9	2.607	-1.413	-0.464	0.379	1.331	50.755	65	.902
10	2.776	-0.768	0.307	1.140	2.037	37.579	57	.978
11	3.167	-1.402	-0.318	0.488	1.307	70.876	56	.087
12	1.850	-1.118	0.212	1.149	2.356	91.428	71	.052
13	2.957	-1.409	-0.362	0.486	1.388	62.737	58	.312
14	1.948	-0.938	0.369	1.198	2.074	74.348	76	.532
15	1.744	-0.631	0.656	1.622	2.830	61.720	70	.749
16	1.208	0.125	1.6889	2.561	3.644	64.742	64	.450
17	1.262	-0.810	0.680	1.634	2.787	77.330	85	.711

a: Item discrimination Index
b_x: Category boundary locations

In Baker's (2001) classification, the item discrimination parameter is moderate for the items between .65 and 1.34, highly discriminating for the items between 1.35-1.69 and very high for items larger than 1.70. According to this, it can be said that item 4, 8, 16 and 17 are moderate, other items are high and very high-level discriminators. These

findings showed that the scale items had adequate to discriminate people with high- and low-level attitudes. The option response functions of the items and item information functions are given in Figure 2 and Figure 3.
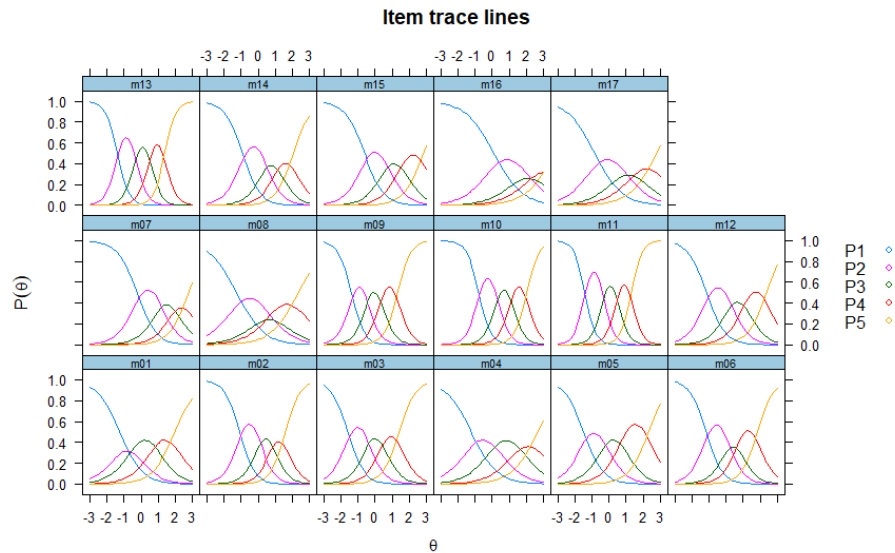

Figure 2. *Option Response Functions for the Items*

The Option Response Functions (ORFs) showed that respondents could not distinguish the categories for some of the items, such as 1, 6, 8, 16, and 17. Some of these items also had a very flat information curve as seen on Figure 3. As a result, a five-point Likert scale did not work for these statements.
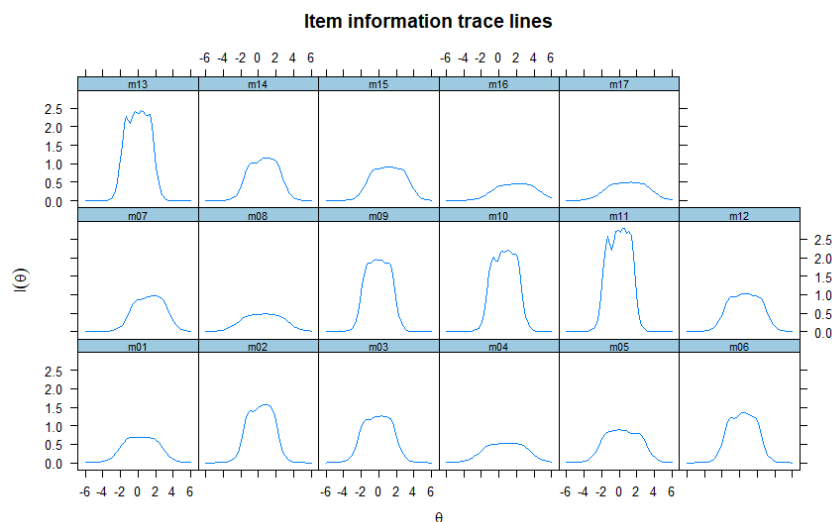

Figure 3. *Item Information Functions*

Item information functions showed that items 4, 8, 16 and 17, which had moderate discrimination parameters, were flatter compared to other items. In other words, these items could not provide information as more as other items for the respondents on the different side of the trait spectrum. Maybe this is not a problem with a homogenous group, but it is otherwise for the studies with heterogenous groups because these items were not capable enough to provide information about respondents. Information and standard error functions of the whole scale are shown in Figure 4.
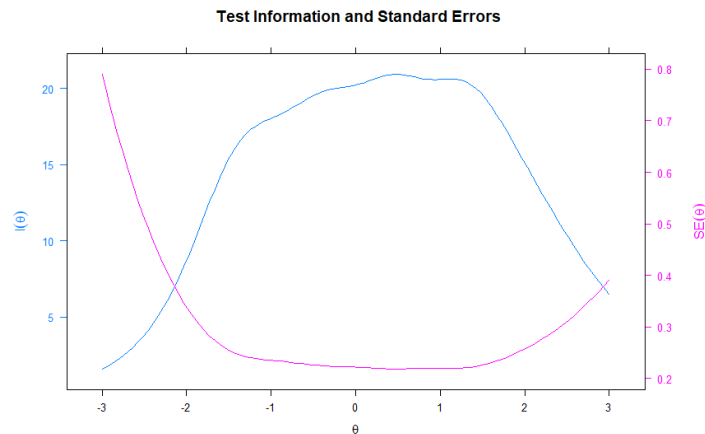
Test Information and Standard Errors

Figure 4. *Test Information and Standard Error Functions*

Test information function in the graph showed that the scale provided more information for individuals with a trait level between -1.5 and 1.5; therefore, it made measurements with fewer errors. In addition, the graph showing how marginal reliability coefficient changes according to the trait level is given in Figure 5.
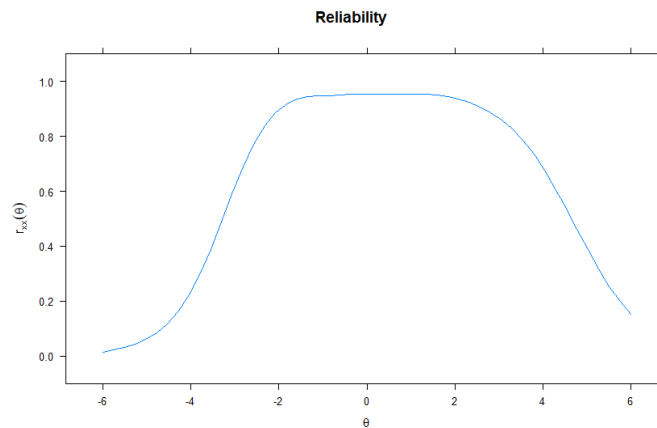
Reliability

Figure 5. *Marginal Reliability Coefficient According to Trait Level*

When the function graph according to the trait level of marginal reliability coefficient was analyzed, the findings showed that the scale made very reliable measurements for individuals whose trait level was between -2 and 3. This finding suggests that this scale may not be the best option for the respondents with very low attitude level. Moreover, the marginal reliability coefficient was estimated, and it was .946, which indicated a very high level of reliability for the scale.

Measurement Invariance and Differential Item Functioning

Measurement invariance among the gender groups was tested to investigate whether the scale measures the same construct for men and women, or not. Table 3 shows the statistics for measurement invariance.

Table 3

Findings on Testing of the Measurement Invariance

Measurement invariance	df	χ^2	RMSEA	CFI	ΔCFI
Configural	208	355.18	.053	.954	
Metric	224	378.76	.052	.952	.002
Scalar	240	407.29	.053	.947	.005
Strict	241	417.32	.054	.944	.003

As seen in Table 3, measurement invariance was met even for the strict level (ΔCFI < .01), which meant that the scale showed invariance between gender groups in the manner of factor structure, factor loadings, indicator intercepts, and indicator error variances. In other words, this scale can be used in gender comparisons in the future. Moreover, DIF test results indicated that only Item 7 showed the possibility to have DIF ($p_{\chi^2_{12}}$ = .005; $p_{\chi^2_{13}}$ = .016; $p_{\chi^2_{23}}$ = .666). However, when the R^2 changes were examined, Item 7 did not seem biased by gender (ΔR^2 < .13). Since the χ^2 statistic was highly depends on the sample size, and ΔR^2 was less than the cut-off value, it could be assumed that none of the items showed DIF by gender.

Discussion, Conclusion and Recommendations

In this study, a 17-item five-point Likert-type unidimensional attitude scale was developed according to the Graded Response Model to investigate university students' attitudes towards pirated content. When the information function of the scale is analyzed, it is seen that the scale provides more information for individuals between the trait level -1.5 and 1.5; however, the scale provides quite a reliable measurement for the individuals with the trait level -2 and 3. This range is compatible with the findings of Demirtasli et al. (2016) and Rubio et al. (2007). In addition, the marginal reliability of the scale is .946. The attitude scale towards pirated content can be used in all kinds of studies in which this trait is planned to be measured. It should be noted that the scale was developed using the Graded Response Model, one of the IRT models. Therefore, evaluation of individuals should be performed with trait

estimations within the context of IRT. The increase in the trait level estimated from the scale means that the attitude towards pirated content has increased. In addition, exploratory factor analysis was applied to test unidimensionality and local independence assumptions while scaling according to IRT. Therefore, the scale can be used based on CTT. Given that the items in the scale do not change according to the CTT and IRT also coincides with the findings of the research conducted by Celen and Aybek (2013) and Yasar (2019).

This scale also shows measurement invariance and DIF-free according to the gender variable of the respondents. In other words, ASPC would be useful for future research that investigates gender differences in attitudes towards pirated content. A new application and item calibrations can be made using only high and very high items with the discrimination parameter in the future research about this subject. Since this study is limited to the Graded Response Model, it would be useful to calibrate the items to other IRT models, such as the Generalized Partial Credit Model and Nominal Response Model. In addition, it is possible to search whether similar trait assumptions can be made using fewer items and computerized adaptive test simulation and live applications which are more applicable with IRT. Yasar and Aybek (2019) reported that response categories of the seven-point Likert scale could not be distinguished by the respondents, which is consistent with the findings of the current study. Thus, future studies could investigate the effects of the response categories to the option response functions.

References

Acilar, A. (2010). Demographic factors affecting freshman students' attitudes towards software piracy: An empirical study. *Issues in Informing Science and Information Technology, 7*. doi: 10.28945/1209

Al-Rafee, S., & Cronan, T. (2006). Digital piracy: Factors that influence attitude toward behavior. *Journal of Business Ethics*, *63*(3). 237-259. doi: 10.1007/s10551-005-1902-9

Baker, F. B. (2001). *The basics of item response theory.* Retrieved December 12, 2019 from http://echo.edres.org:8080/irt/baker/final.pdf

Baykul, Y. (2010). *Egitimde ve psikolojide olcme: Klasik test teorisi ve uygulamasi. [Measurement in education and psychology: Classical test theory and application].* Ankara: OSYM Yayinları.

BSA. (2013). *Yazilim korsanligi nedir? [What is software piracy?]* Retrieved from http://ww2.bsa.org/country/Anti-Piracy/What-is-Software-Piracy.aspx

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. doi: 10.18637/jss.v048.i06

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. doi: 10.1080/ 10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Choi, S.W. (2016). lordif: Logistic ordinal regression differential item functioning using IRT. R package version 0.3-3. Retrieved from https://CRAN.R-project.org/ package=lordif

Celen, U., & Aybek, E. C. (2013). Ogrenci basarisinin ogretmen yapimi bir test ile klasik test kurami ve madde tepki kurami yontemleriyle elde edilen puanlara gore karsilastirilmasi [Comparison of student achievement with a teacher-made test according to scores obtained by classical test theory and item response theory methods]. *Egitimde ve Psikolojide Olcme ve Degerlendirme Dergisi*, 4(2), 64-75. Retrieved from https://dergipark.org.tr/tr/download/article-file/ 65958

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

Demirtasli, N., Yalcin, S., & Ayan, C. (2016). The development of IRT based attitude scale towards educational measurement course. *Journal of Measurement and Evaluation in Education and Psychology, 7*(1). 133-144. doi: 10.21031/epod.4380

Dilmperi, A., King, T., & Dennis, C. (2017). Toward a Framework for Identifying Attitudes and Intentions to Music Acquisition from Legal and Illegal Channels. *Psychology & Marketing, 34*(4), 428-447. doi: 10.1002/mar.20998

Field, A. (2009). *Discovering statistics using SPSS and sex and drugs and rock 'n' roll.* London: SAGE.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory.* New York: Springer Science+Business Media

International Federation of the Phonographic Industry [IFPI]. (2013). *What is piracy?* Retrieved from http://www.ifpi.org/content/section_views/what_is_ piracy.html

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). *semTools: Useful tools for structural equation modeling.* R package version 0.5-3. Retrieved from https://CRAN.R-project.org/package=semTools

LaHuis, D. M., Clark, P., & O'Brien, E. (2011). An examination of item response theory item fit indices for the graded response model. *Organizational Research Methods, 14*(1), 10-23. doi: 10.1177/1094428109350930

MU-YAP. (2013). *Muzik korsanligi nedir? [What is music piracy?].* Retrieved from http://mu-yap.org/getdata.asp?pagename=korsan_nedir

Okurame, D., & Ogunfowora, A. (2011). Impact of gender and opportunity recognition on attitude to piracy of computer industry products. *Gender & Behavior, 9*(1). doi: 10.4314/gab.v9i1.67449

Pastore, A., & Cesareo, L. (2014). Consumers' attitude and behavior towards online music piracy and subscription-based services. *Journal of Consumer Marketing, 31*(6/7), 515-525. doi: 10.1108/JCM-07-2014-1070

R Core Team. (2013). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rubio, V.J., Aguado, D., Hontangas, P.M., and Hernández, M. (2007). Psychometric properties of an emotional adjustment measure. *European Journal of Psychological Assessment, 23*(1). 39-46. doi: 10.1027/1015-5759.23.1.39

Samejima, F. (1996). *Polychotomous responses and the test score.* Tennessee: The University of Tennessee.

Sheehan, B., Tsao, J., & Pokrywczynski, J. (2012). Stop the music! How advertising can help stop college students from downloading music illegally. *Journal of Advertising Research, 52*(3), 309-321. doi: 10.2501/JAR-52-3-309-321

Smits, N, Zitman, F.G., Cuijpers, P., den Hollander-Gijsman, M.E., & Carlier I.V. (2012). A proof of principle for using adaptive testing in routine outcome monitoring: The efficiency of the mood and anxiety symptoms questionnaire-anhedonic depression CAT. *BMC Medical Research Methodology, 12*(1). doi: 10.1186/1471-2288-12-4

Thurstone, L.L. (1928). Attitudes can be measured. *The American Journal of Sociology, 33*(4). doi: 10.1086/214483

Yasar, M. (2019). Development of a "perceived stress scale" based on classical test theory and graded response model. *International Journal of Assessment Tools in Education, 6*(3). 522-538. doi: 10.21449/ijate.626053

Yasar, M., & Aybek, E.C. (2019). A resilience scale development for university students: Validity and reliability study based on item response theory. *Elementary Education Online, 18*(4). 1687-1699. doi: 10.17051/ilkonline.2019.635031

Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187–213. doi: 10.1111/j.1745-3984.1993.tb00423.x.

Yoon, C. (2010). Theory of planned behavior and ethics theory in digital piracy: An integrated model. *Journal of Business Ethics, 100,* 405-417. doi: 10.1007/s10551-010-0687-7

Yu, S. (2012). College students' justification for digital piracy: A mixed methods study. *Journal of Mixed Methods Research, 6*(4), 364-378. doi: 10.1177/1558689812451790

Zumbo, B. (1999). *A handbook on the theory and methods of differential item functioning (DIF).* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defence.

Korsan İçeriğe Yönelik Tutum: Aşamalı Tepki Modeli Temelinde Bir Ölçek Geliştirme Çalışması

Atıf:

Özet

Problem Durumu: Korsan içeriğe yönelik tutumun belirlenmesi amacıyla yapılan çalışmalarda, veri toplama aracı olarak genellikle anketlerin kullanıldığı görülmüştür. Bu çalışmalarda herhangi bir ölçek geliştirme yoluna gidilmemiş, dolayısıyla bir geçerlilik ve güvenilirlik çalışması da yapılmamıştır. Bu nedenle, bu araçlardan toplam puan elde edilmesi ve toplam puanlar üzerinden işlemlerin yapılması mümkün değildir. Bununla birlikte Klasik Test Kuramı temelinde geliştirilen çevrimiçi korsan içeriğe ve çevrimiçi müzik korsanlığına yönelik ölçekler bulunmaktadır. Birey davranışlarını bilmenin, korsan karşıtı kuruluşlar ya da devletler tarafından izlenecek politikalarda önemli rol oynayacağı düşünüldüğünde, bireylerin korsan içeriğe yönelik tutumlarını belirlemeye yönelik bir ölçeğin geliştirilmesine ihtiyaç duyulmaktadır.

Araştırmanın Amacı: Araştırmanın amacı, üniversite öğrencilerinin korsan içeriğe yönelik tutumlarını belirlemeyi amaçlayan, ölçme sonuçlarının güvenilir ve geçerli olduğu bir ölçme aracını Madde Tepki Kuramı temelinde geliştirmektir.

Araştırmanın Yöntemi: Bu araştırmada, 494 üniversite öğrencisine ulaşılmıştır. Ancak geçersiz formların ve uç değerlerin veri setinden çıkarılmasının ardından, 89 kadın, 346 erkek ve cinsiyetini belirtmeyen 3 katılımcı olmak üzere toplam 438 üniversite öğrencisi araştırma grubunu oluşturmuştur.

Korsan İçeriğe Yönelik Tutum Ölçeği (KİYTÖ) geliştirilmeye başlamadan önce test geliştirme planı oluşturulmuş ve bu doğrultuda ölçek ile ölçülmesi amaçlanan yapıya dair literatür taramasıyla psikolojik yapı tanımlanmıştır. Daha sonra, ölçek maddelerinin yazılması için literatür taramasının yanında, araştırma grubunun dışında yer alan 60 üniversite öğrencisine korsan içerikle ilgili kompozisyonlar yazdırılmış ve toplam 53 madde oluşturulmuştur. Oluşturulan maddeler üç ölçme ve

değerlendirme uzmanı tarafından incelenmiş ve tutum ifadesi olmadığı, ölçülmek istenilen özelliği ölçmediği, ifadesinin belirsiz olduğu düşünülen maddeler çıkarılarak toplam 17 maddelik ölçek deneme formu oluşturulmuştur. Ölçek deneme formu Hiç katılmıyorum, Katılmıyorum, Kısmen katılıyorum kısmen katılmıyorum, Katılıyorum ve Tamamen katılıyorum olmak üzere beşli Likert tipinde yanıtlama seçeneklerine sahip; 9'u olumsuz, 8'i olumlu toplam 17 maddeden oluşmuştur.

Ölçeğin yapı geçerliği çalışması sırasında temel bileşenler analizi (principal component analysis) yerine, temel eksen ayırma (principal axis factoring) tekniği ile açımlayıcı faktör analizi yapılmıştır. Bunun nedeni, araştırmanın matematiksel özetlemelerden ziyade gizil bir yapıyı ortaya çıkarma çabasıdır. Birden fazla faktörün bulunması durumunda eksen döndürme tekniği olarak dik döndürme tekniklerinden varimax tekniği kullanılmasına karar verilmiştir. Dik döndürme yönteminin tercih edilme nedeni ise analiz sonuçlarının mevcut verilerle en iyi şekilde uyum göstermesi değil, genellenebilirliğinin amaçlanmasıdır. Faktör sayısına karar verilirken hem faktör özdeğerleri arasındaki ilişkiler hem de yamaç birikinti (scree plot) grafiği dikkate alınmıştır.

Araştırmanın Bulguları: Madde Tepki Kuramı varsayımlarından tek boyutluluk varsayımının karşılanıp karşılanmadığını belirlemek amacıyla açımlayıcı faktör analizi yapılmış ve maddelerin tek bir başat faktör altında yer alıp almadığı incelenmiştir. Veri setinin açımlayıcı faktör analizi için uygun olup olmadığını belirlemek için Kaiser-Meyer-Olkin (KMO) ve Bartlett küresellik testi sonuçları incelenmiştir. KMO değerinin .92 ve Bartlett küresellik testi sonucunun manidar olması ($\chi^2 = 3559.897$; p< .05) veri setinin açımlayıcı faktör analizine uygun olduğunu göstermiştir. Açımlayıcı faktör analizinde faktör çıkarma tekniği olarak temel eksen ayırma tekniği kullanılmış ve bunun sonucunda 17 maddenin tek bir faktör altında toplandığı görülmüştür. Yerel bağımsızlığın sağlanıp sağlanmadığını belirlemek içinse Q3 istatistikleri hesaplanmıştır. Buna göre her bir madde çifti için artık değerler arası korelasyon değerleri hesaplanmış ve 2-3; 2-7; 11-13; 11-16; 12-14 ve 15-16 maddeleri arasında orta düzeyde bağımlılık olduğu bulunmuştur. Yüksek düzeyde bağımlılık gösteren herhangi bir maddeye ise rastlanmamıştır. Bunun yanında kestirilen yetenek düzeylerinin güvenilirliğini belirlemek için hesaplanan marjinal güvenilirlik katsayısı ise .946 olarak elde edilmiştir. Maddelerin model ile uyum gösterip göstermediğinin belirlenmesinde S-χ2 istatistiğinden yararlanılmıştır. Buna göre tüm maddelerin model ile uyum gösterdiği (p>.05) sonucuna ulaşılmıştır. Ayırt edicilik parametresi .65 ile 1.34 arası olan maddeler için orta düzeyde ayırt eden, 1.35-1.69 arası olan maddeler için yüksek düzeyde ayırt eden ve 1.70'ten büyük maddeler için ise çok yüksek düzeyde ayırt eden madde nitelemesi yapılmaktadır. Buna göre, 4, 8, 16 ve 17. maddelerin orta düzeyde, diğer maddelerin ise yüksek ve çok yüksek düzeyde ayırt edici maddeler olduğu sonucuna ulaşılmıştır. Test bilgi fonksiyonu incelendiğinde, ölçeğin -1.5 ile 1.5 arasında özellik düzeyine sahip bireyler için daha çok bilgi sağladığı, dolayısıyla bu aralıktaki bireyler için daha az hata ile ölçme yaptığı görülmektedir. Marjinal güvenilirlik katsayısının özellik düzeyine göre fonksiyon grafiği incelendiğinde, ölçeğin özellik düzeyi -2 ile 3 aralığındaki bireyler için güvenirliği yüksek ölçmeler yaptığı sonucuna ulaşılmıştır.

Araştırmanın Sonuçları ve Önerileri: Araştırma sonunda, Aşamalı Tepki Modeli (Graded Response Model) temelinde, korsan içeriğe yönelik bir tutum ölçeği geliştirilmiştir. Öte yandan, ölçeğin özellikle -1.5 ve 1.5 özellik düzeyindeki bireyler için daha çok bilgi sağladığı sonucuna ulaşılmıştır. Ölçekten elde edilen puanların artması, öğrencilerin korsan içeriğe yönelik olumlu tutumlarının arttığı, azalması ise ilgili tutumlarının azaldığı şeklinde yorumlanabilir. İleride bu konuda yapılacak olan çalışmalarda ayırt edicilik parametresi yalnızca yüksek ve çok yüksek maddeler kullanılarak yeni bir uygulama ile madde kalibrasyonları yapılabilir. Bu araştırma Aşamalı Tepki Modeli ile sınırlı olduğundan, korsan içeriğe yönelik tutum ölçeğinin Genelleştirilmiş Kısmi Puan Modeli, Sınıflamalı Tepki Modeli gibi farklı MTK modelleriyle de ölçeklenmesi yararlı olacaktır. Ayrıca MTK ile daha uygulanabilir olan Bilgisayarda Bireyselleştirilmiş Test simülasyonu ve canlı uygulamaları ile daha az madde kullanılarak benzer özellik kestirimlerinin yapılıp yapılamayacağı araştırılabilir.