International Journal of Assessment Tools in Education
2020, Vol. 7, No. 1, 18–29
https://dx.doi.org/10.21449/ijate.629584
Published at http://www.ijate.net http://dergipark.org.tr/en/pub/ijate *Research Article*

Use of Item Response Theory to Validate Cyberbullying Sensibility Scale for University Students

Osman Tolga Arıcak [1], **Akif Avcu** [2,*], **Feyza Topçu** [1], **Merve Gülçin Tutlu** [1]

[1]Department of Psychology, Hasan Kalyoncu University, Gaziantep
[2]Department of Educational Sciences, Marmara University, Istanbul

ARTICLE HISTORY

Received: 04 October 2019
Revised: 03 December 2019
Accepted: 06 Febuary 2020

KEYWORDS

Cyberbullying sensibility,
Test validation,
Item response theory,
Graded response model,
Item selection

Abstract: A thirteen-item cyberbullying sensibility scale (CSS), developed by Tanrıkulu, Kınay, and Arıcak (2013) and extensively used by researchers, was used to measure the cyberbullying sensibility levels of high school students. Unlike other similar concepts, such as cyberbullying and cyber victimization, there are no scales developed to measure the cyberbullying sensibility among university students. In this study, the data obtained from 727 university students were analyzed based on item response theory (IRT) techniques, and psychometric evidences were obtained to evaluate whether it is appropriate to use the scale on the university students. Accordingly, a parameterization of CSS items was performed by using the graded response model. Using the discrimination parameters and item fit statistics, some items were removed from the original scale and a seven-item CSS version was developed since preliminary exploratory and confirmatory factor analyses provide inadequate evidence for the validity of a one-dimensional structure of cyberbullying sensibility. However, an IRT-based item removal process yielded an acceptable improvement. In this way, despite the six items being removed from the original CSS form, the scale retained 64% of the information it provided. The reliability values computed based on the classical approach and IRT were above .8 after the item elimination process with only a minor drop. With the validation process, the CSS will be a valuable measurement tool to determine the level of cyberbullying sensibility among university students and allow academicians to conduct research with this population.

1. INTRODUCTION

Today, young people who use information technologies are under extreme risk of cyberbullying in cyberspace's unknown and virtual social relations (Willoughby, 2018). They may be engaging in bullying, be exposed to bullying, or be bystanders (Gahagan, Vaterlaus, & Frost, 2016). A noteworthy concept in prevention of bullying in cyberspace is the cyberbullying sensibility. Studies show that young people not only intentionally hurt others, but also can bully others just for fun (Arıcak, 2015; Tolia, 2016). This is an important finding that emphasizes the lack of knowledge and awareness of the consequences of such actions. The cyber-bullying sensibility is defined as the awareness of young people about cyberbullying behaviors while

CONTACT: Akif Avcu avcuakif@gmail.com Marmara University, Atatürk Education Faculty, İstanbul, Turkey

ISSN-e: 2148-7456 /© IJATE 2020

using electronic media and how sensitive they are toward these kinds of behaviors (Tanrıkulu, Kınay, & Arıcak, 2015). Studies aimed at preventing cyberbullying (Gaffney, Farrington, Espelage, & Ttofi, 2018) and increasing the sensibility to cyberbullying (Nedim Bal & Kahraman, 2015; Tanrıkulu, Kınay & Arıcak, 2015) have begun to take their place in the literature. This trend in the literature requires the use of instruments that measure the sensibility to cyberbullying behaviors. For this aim "Cyberbullying Sensibility Scale" (CSS), developed by Tanrıkulu, Kınay and Arıcak in 2013 to measure the cyberbullying sensibility of adolescents, has been used in various studies in the last six years (i.e., Aktan & Çakmak, 2015; Baştak & Altınova, 2015; Doğan, Cansu, & Şahin, 2016).

IRT (item response theory) is an important psychometric approach used in the processes to obtain valid measurement instruments. Its use has become widespread among test developers since this method can solve many measurement difficulties encountered during test development process and provide richer output (Samejima, 1968; Embretson, 1996). For this reason, the validation of the CSS for university students was carried out by taking advantage of IRT.

The most salient difference between the Classical Test Theory (CTT) and IRT is that the CTT assumes equal measurement accuracy across all test takers, regardless of their ability levels. However, in IRT, the measurement accuracy depends on the level of the latent trait being measured. This leads to a differentiation between results obtained from CTT and IRT. When model-data fit is achieved, IRT provides the test information functions (TIF) (the amount of the information the test provides to the users) and the amount of error for different ability levels (Hambleton et al., 2000).

As stated by Hambleton et al. (1991), IRT models have two basic assumptions provided that the measured property is one-dimensional: unidimensionality and local dependence. The first assumption of unidimensionality means that only one trait is being measured by a set of items composing the test. It requires the presence of a dominant factor explaining most of the variability on test scores. In other words, the covariance between items can be explained by the single dimension. Hattie (1985) recommended that the unidimensionality could be tested with investigating eigenvalues and variability explained by the first factor based on exploratory factor analysis (EFA) testing one dimensional factor analysis via confirmatory factor analysis (CFA). Another IRT assumption is local independence. According to this assumption, responses to an item should not be statistically related to each other, even after the latent trait being measured is kept statistically constant.

Another IRT assumption is local independence. According to this assumption, responses to an item should not be statistically related to each other, after the latent trait being measured is kept statistically constant. It implies that, an observed responses must not be affected by any unrelated factors other than the ability levels of participants. Different statistics developed so far to investigate whether or not local independence assumption holds. Most commonly preferred statistics are χ^2 statistics, G2 statistics (Chen & Thissen, 1997) and Q3 statistics (Yen, 1984). For the current study, only Q3 statistics were taken into account in order to determine whether Local dependence (LD) was present or not. Even though there is no consensus on the cut off value of Q3 statistics, the value of 0.3 were generally considered as an evidence for the existence of LD.

Many different models have been developed to analyze Likert-type items with more than two response options also known as polytomous items. Although these polytomous response models differ among themselves in terms of parametrization, they all include the specification of a location and slope parameter (and the characteristic curve accordingly) for each response category (Thissen & Steinberg, 1986). The Graded Response Model (GRM) (Samejima, 1968) is a polytomous IRT model developed for item responses characterized by graded categories.

The GRM and is considered to be the generalization of two parameter logistics models (Keller, 2005). The model is particularly suitable for use with Likert type items and has beed preferred in different studies (i.e. Rubio et al., 2007; Mielenz et al., 2010). Even though there are some other alternative polytomous models available in the literature (see Hambleton & Swaminathan, 1985) GRM was preferred for the current study.

The item level fit statistics could also be obtained by utilizing IRT based approach and was evaluated with polytomous extension of S-X^2 item-fit index (Orlando & Thissen, 2000). This index is Chi-square based and uses a significance test. Generally, *p* values lower than .05 were considered a poor item fit.

Another important advantage of the IRT is the provision of item and test information. In IRT terminology, the term information implies the amount of accuracy of the measurement and closely related to reliability. For a two-parameter model, the item information is determined as the function of the item discrimination and the item location parameters in each value of the ability parameter. The item information function shows the contribution of each item to the measurement of the latent trait being measured. Items with more discrimination power contribute more to the accuracy of measurement (Hambleton et al., 2000).

Given these advantages cited above, the use of IRT in test development/revision processes provides advantages that cannot be achieved with classical test theory. Accordingly, the number of studies using IRT for test development, test revision and obtaining shorter versions of available ones increased in the last decade (i.e. Zanon, Hutz, Yoo & Hambleton, 2016; Istiyono et al., 2019; Bilker et al., 2012). In the light of this fact, the revision of CSS using IRT based approach is the main purpose of this study.

2. METHOD

2.1. Study Group

The participants in this study were 727 university students. The average age of students was 22.03 (SD=22.03, ranged between 18-26 years), and the majority, 462, were female (63.4%) and the rest, 266, were male (36.6%). The participants were selected with convenient sampling among the students studying in various faculties of a private university. Since IRT studies are model-based, large sample size is of great importance for the accuracy of the measurement. For models with more parameters estimated, Tsutakawa and Johnson (1990) stated that a sample size of 500 would be sufficient. For this reason, participation in the research was on a voluntary basis, and the sample size was intentionally kept high, above the size recommended by Tsutakawa and Johnson (1990).

2.2. Measurement Instrument

The Cyberbullying Sensibility Scale (CSS) was developed by Tanrıkulu, Kınay & Arıcak in 2013. The scale development process was conducted in Istanbul metropolitan area with 663 high school students. For construct and construct validity, both the EFA and the CFA were carried out. The results of the EFA showed that the scale has a one-dimensional structure, explaining 47% of the total variance with factor loadings varying between .61 and .76. The CFA results further confirmed the one-dimensional structure of the scale (χ^2/df=3.22, RMSEA=.082). The loadings coefficients varied between .31 and .65. The Cronbach alpha coefficient was estimated as .90 and the test-retest reliability as .63.

2.3. Procedures

The participants were asked to answer the items of CSS and a demographic information form in their own classes. The students were informed that participation in the research was voluntary and that any information they provided would be kept confidential. The students were asked to read the questions carefully and fill in the scale to reflect their views. Data collection was

performed in a single session for each class.

2.4. Analysis

To test the unidimensionality of the data, a confirmatory factor analysis was performed by utilizing MPLUS 6 (Muthén & Muthén,1998-2012). The fit of the model was evaluated χ^2 statistic, Root Mean Square Approximation (RMSEA), standardized root mean squared residual (SRMR) and confirmatory fit index (CFI) and Tucker-Lewis index (TLI) indexes were used. In addition, unidimensionality was also evaluated by appling an exploratory factor analysis. For this analysis, SPSS 21 was utilized. In addition, local independence assumption was evaluated by Q3 statistics (Yen, 1984). After checking the assumptions. The GRM (Samejima, 1968) was used for IRT based item calibration. After obtaining item parameters, item fit statistics were computed with S-X^2 index (Orlando & Thissen, 2000). IRT based factor analysis was performed with a *mirt* package (Chalmers, 2012) in R program (R core team, 2017). In addition, the item level fit statistic and information functions for the test and items were computed with the same package.

3. RESULT / FINDINGS

3.1. Testing IRT assumptions

The confirmatory factor analysis (CFA) results showed that a one-dimensional model was not confirmed at acceptable level: χ^2 (N=727, df=65) =492.62, $p<0.001$, CFI =0.860 TLI =0.832, RMSEA=.95, %95C.I. = [0.087-0.102], and SRMR =0.053. Factor loadings varied between 0.44 and 0.70 at $p<0.001$ significance level. In addition, an exploratory factor analysis (EFA) was conducted as a supplemental. The results showed that two factors were extracted with eigenvalues greater than one. The first factor explained 39.95% of total variance while the second factor explained 8.91%. Even though the CFA results did not confirm the one factor structure of the CSS, the EFA results provided evidence that the original unidimensional structure of the CSS was present because, as stated earlier, Hattie (1985) stated that 20% or more variability on the first factor is a proof for the presence of a dominant factor explaining most of the instrument scores. At this point, IRT based analyses were performed to obtain more acceptable results to keep original one-dimensional structure of the CSS for university students. We achieved this by investigating the information contribution of each item and item level fit statistics.

The LD assumption was tested by computing the Q3 statistics. The results showed that only one item pair had a Q3 value greater than 0.3 (item 1 - item 2). At this point, one of the items had to be eliminated from the scale. We eliminated item 1 because it was not only locally dependent with item 2 and it didn't satisfy the criterion related to the amount of information it had (see below). In addition, the LD was tested with a final 7-item version of the scale, and no item pairs were found with LD.

3.2. Fitting the Graded Response Model

As previously mentioned, the CSS has three graded response options. Hence, the GRM was preferred for model fitting. As a result of the GRM estimations, one discrimination and two threshold parameters were obtained. The analysis showed that the fit values of the single-factor GRM model were at an acceptable level: CFI =0.942 TLI =0.954, RMSEA=.074, and SRMR =0.067.

Table 1 lists item parameters, information results, and S-X^2 statistics to evaluate item fit. The factor loadings for the CSS items ranged from .52 to .72. In addition, the communality values ranged from .33 to .63. The results provided evidence that the items are fairly different in terms of the amount of variation with a common factor. The discrimination parameters ranged from 1.02 to 2.44. The contribution of items to total information varied between 1.49 and 4.19, and

the percentage of the contribution to the total test information varied between 4.94% and 11.11%. This finding shows that the contribution of items to the model shows a significant variance. The difficulty parameters of the CSS vary between -0.62 and 4.02. This finding shows that the CSS is not useful enough for identifying individuals with low levels of cyberbullying sensibility but was a scale more suitable for identifying average-to-high-level individuals. The item fit was also evaluated by using S-X^2 statistics. The results showed that three items (item 7, item 12, and item 13) had p values less than .05, which suggest that the items did not fit the model well.

3.3. Item Selection for the CSS for University Students

The items were selected from the 13 item CSS to increase the model fit index, contributing to the validation process of the scale for university students. Two different criteria were used in this process. First, the amount total item information was obtained by adding of the information values of each item. Later, total item value was divided by the number of items in CSS. In this way, average information value was obtained. The items with above-average contribution to total information were determined, and these items were kept in the new CSS form. This operation was carried out only once. Secondly, the items with a poor level of item fit statistics (items with p values corresponding to S-X^2 statistics below .05) were also eliminated from the new form of the CSS. This process continued until no poor fit items remained. As seen in Table 1, five items below the average were removed from the scale. Three of these items also had a poor fit ($p<.05$). Thus, the second criterion was not used to eliminate items at this stage. After removing these five items from the scale, the GRM model was repeated with the remaining eight items, the item fit statistics were examined, and each item with a poor fit level was determined and eliminated (item 1). When the GRM model was repeated with the remaining seven items, it was found that there were no items to be eliminated according to this criterion.

Table 1. Initial item loadings, communalities, four parameters, fit statistics, and information

	Factor analysis		Item Parameter			Item Fit			Test info=37.74 (M=2.90)	
	F	h2	a1	d1	d2	S-X^2	df (S-X^2)	*p* (S-X^2)	Item info.	%
item 1	0.77	0.59	2.04	0.77	3.94	29.84	25	0.230	3.68	9.74
item2	0.79	0.62	2.19	0.95	4.07	17.45	25	0.865	3.91	10.35
item3	0.52	0.27	1.02	0.23	1.95	31.11	25	0.185	**1.49**	3.96
item4	0.73	0.53	1.82	0.32	2.87	39.14	29	0.099	3.02	8.00
item5	0.75	0.56	1.91	0.24	2.65	24.95	29	0.681	3.14	8.32
item6	0.73	0.53	1.80	0.47	3.07	32.79	29	0.286	3.05	8.08
item7	0.61	0.37	1.31	0.99	3.11	58.90	33	**0.004****	**2.11**	5.60
item8	0.58	0.33	1.20	1.10	3.00	39.46	34	0.239	**1.87**	4.94
item9	0.82	0.67	2.44	0.07	2.95	30.58	26	0.244	4.19	11.11
item10	0.72	0.52	1.77	-0.49	2.21	33.77	28	0.209	2.96	7.84
item11	0.79	0.63	2.20	0.18	2.89	24.35	27	0.611	3.62	9.60
item12	0.65	0.43	1.46	-0.62	2.01	48.03	31	**0.026***	**2.38**	6.29
item13	0.70	0.48	1.65	1.03	2.56	39.53	22	**0.012***	**2.33**	6.17

Note: *$p<0.05$; **$p<0.01$; h2=communality; bolded results are pointing the items with below average information contribution and poorly fitting items at first stage investigation.

3.4. Fitting the Graded Response Model for Seven-item CSS Form

The results for the GRM fitted to the seven-item CSS form are provided at Table 2. Factor loadings ranged from .71 to .85, and communalities ranged from .50 to .72. When compared with the 13-item version, it was seen that the variance of items with common factor showed less variability. Item discrimination values of seven items varied between 1.71 to 2.74. Again,

items in the seven-item CSS form are more homogenous in terms of their discrimination powers. In addition, the results showed that the threshold parameters ranged from .52 to 3.71. In addition, for seven item version, S-X^2 statistics and their corresponding p values suggested that all items were fitted to the GRM model at an acceptable level ($p>.05$). The seven-item version retained 64% of the total information provided by the original CSS, and the average contribution of each item to total test information increased to 18%.

Table 2. Final item loadings, communalities, parameters, fit statistics and information for 7-item CSS.

	Factor analysis		Item Parameter			Item Fit			Test info=24.00 (M=3.43)	
	F	h2	a1	d1	d2	S-X^2	df(S-X^2)	p(S-X^2)	Value	%
item2	0.73	0.54	1.84	3.71	0.83	18.56	16	0.292	3.23	13.47
item4	0.71	0.50	1.71	2.80	0.29	18.34	17	0.367	2.89	12.04
item5	0.74	0.55	1.88	2.65	0.23	18.85	16	0.277	3.15	13.13
item6	0.72	0.52	1.76	3.06	0.45	18.72	16	0.284	3.01	12.54
item9	0.85	0.72	2.74	3.20	0.06	17.82	14	0.215	4.92	20.48
item10	0.75	0.56	1.91	2.30	-0.52	23.20	15	0.080	3.33	13.89
item11	0.76	0.59	2.02	2.76	1.50	10.06	16	0.863	3.47	14.44

Not: h2=communality

3.5. Comparison of the 13-item CSS Form and the Seven-item CSS Form

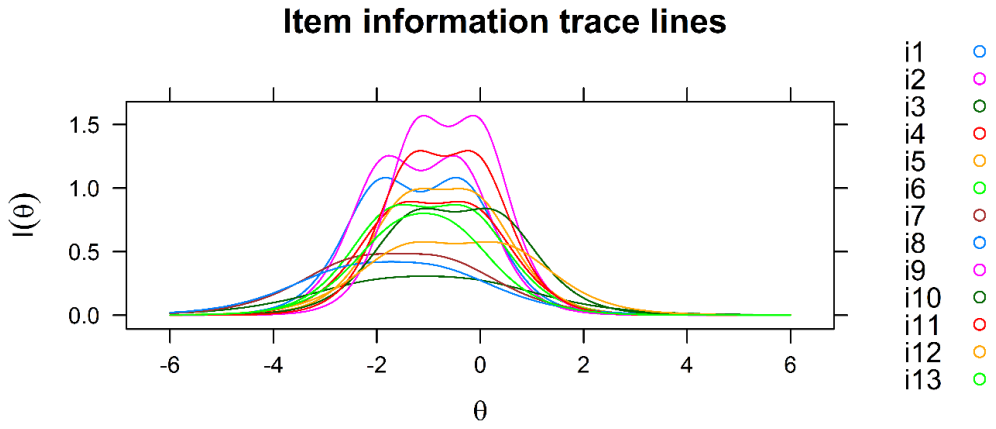
To compare both forms, the CFA analysis was repeated with a seven-item version of the CSS. The results showed that a one factor solution was confirmed with acceptable fit indices: $\chi^2(2) = 101.05$, df = 14, $\chi^2(2)$ / df = 7.21 CFI = 0.943, TLI = 0.914. RMSEA = 0.092(95% CI= 0.076-0.110), SRMR = 0.038. As presented above, the fit indices were not at an acceptable level for the 13-item version. It was clearly seen that removing items contributes to model data fit. The IRT based factor analysis was also computed by fitting the GRM model. The analysis showed that the fit values of the single-factor GRM model were at an acceptable level: CFI =0.976.TLI =0.952. RMSEA=.072. and SRMR =0.055. As compared to the GRM model fitted with 13-items, the seven-item version provides better IRT based fit indices for a one factor solution.

Cronbach alpha coefficients were also computed for both versions in order to see how the internal consistency of the scale was affected by reducing the number of the scale. While the alpha value was .87 for the 13-item version, it was found that the value for the seven-item version was .83, showing a minimal drop of internal consistency after reducing to six items. The IRT-based empirical reliability values were also compared between both forms. The results were similar for both forms (.87 for the 13-item version vs .81 for the seven-item form). The correlation between both forms using were also computed using their total scores. It is .95 ($p<0.001$), indicating that the seven-item version could be used for same purposes as the 13-item version of the CSS.

The two different versions of the scale were compared based on the item and test information functions. The item information functions of both versions are presented in Figures 1a and b respectively. As depicted in Figures 1a and b, both versions of the CSS scale contain items with varying information levels, which contributes to total test information, while most of the less contributing items were eliminated in the seven-item version. In addition, while the information they contribute show variation, the items are slightly homogenous in terms of the location where they provide their highest information. The range of the ability both versions provide is narrow to some extent but similar to each other while it could be inferred that both forms provided accurate measurement within the same range of cyberbullying sensibility levels in terms of the range of the information they provide (see Figure 2).

In the seven-item version, the two items with relatively less information were retained in the test because they could possibly provide information at the level of the ability range. All in all, we can infer that the item removing process mostly eliminates item with lower information except for two items that provided information at different locations.

a



b

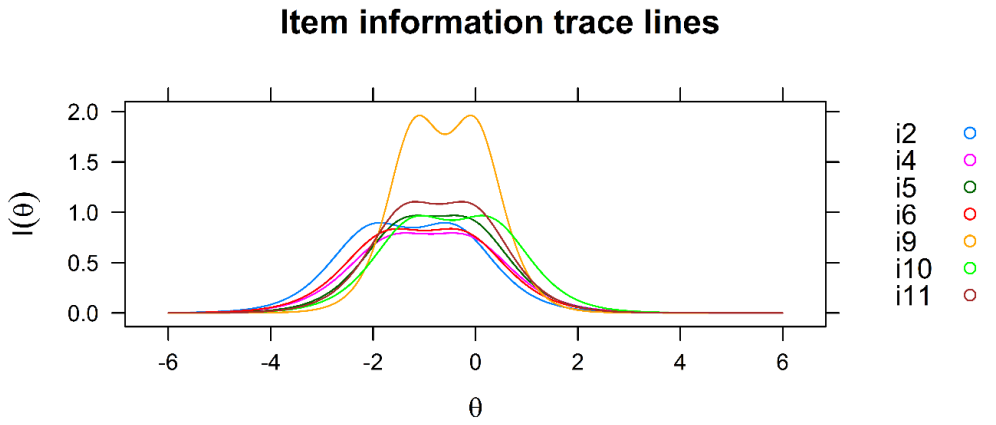
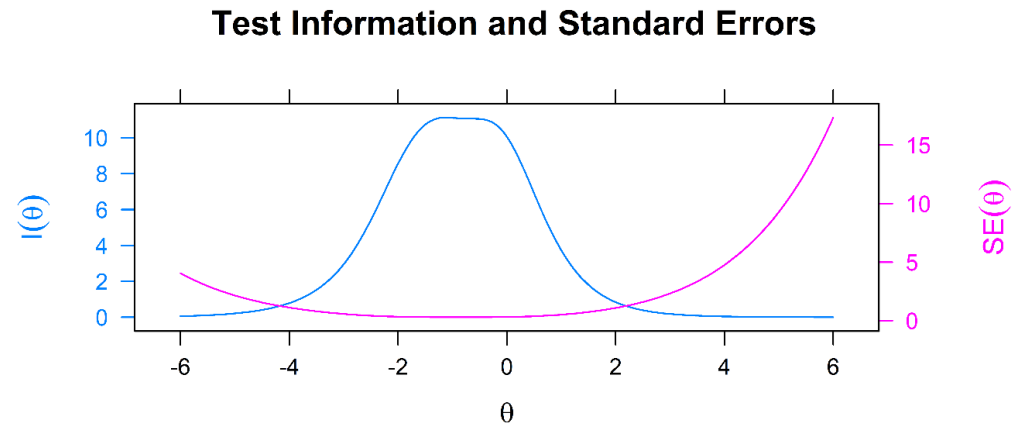


Figure 1. (a) Item Information Curves for thirteen item version of CSS. (b) Item Information Curves for seven items version of CSS. Note: Numbers in shorter version indicate the item number after items removed from original version of CSS

Figures 2a and b represent test information functions and a standard error of measurement of the 13-item and the seven-item versions of the CSS. As shown in Figure 2, both forms are similar in terms of the maximum information they provide across the ability spectrum. As expected, in terms of the information and precision of items, the 13-item version is better because it contains more items while the drop in information (and increase in the standard error) is not comparable with the proportion of the reduced items. The percentage of the remaining items were 54% while 65% of the information still retained.

a



b

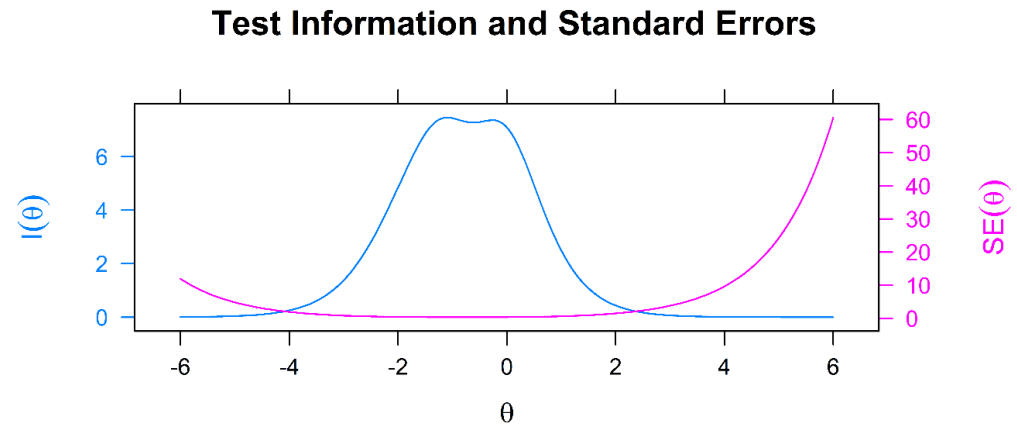


Figure 2. (a)Total test information for the initial 13-item CSS. (b) Total test information function for the seven-item CSS scale.

4. DISCUSSION and CONCLUSION

The primary aim of this research was to validate the 13-item Cyberbullying Sensibility Scale developed by Tanrıkulu, Kınay and Arıcak (2013) for university students. In this respect, item response theory (IRT) was used in order to obtain findings related to the psychometric properties of the scale. IRT based techniques were used to remove some items that interfered with the one-dimensional structure andprovided below-average information. We expect that the IRT based item removal process would not adversely affect the reliability of psychometric properties of the scale and the amount of information it provided. Moreover, we hoped that its usefulness would increase thanks to the shortening of the scale. Shorter scales, in addition to increasing the usefulness in practice by enabling the addition of more variables to the research, expand the nomological network of cyberbullying to other psychological constructs. Finally, this validation study made the measuring of the cyberbullying sensibility among university students possible for interested researchers. Although there are scales of cyberbullying and cyber victimization that could be used with university students, there is a lack of a standardized scale of cyberbullying sensibility suitable for university students. This study will contribute to filling this gap in the literature.

The discrimination parameters obtained from the IRT analyzes showed that the seven-item cyberbullying sensibility scale validated for the university students was effective in distinguishing students with low and high levels of cyberbullying sensibility. On the other hand,

when the IRT-based difficulty parameters were examined, it was seen that the scale provided more accurate measurements for average-to-high-level individuals.

The item removal process was conducted based on two criteria: (a) removing items with below-average contribution to total test information, (b) removing items with poor item fit statistics. As a result of the IRT analysis performed with 13 items in the original form, the fit statistics of two items were found to be less than $p<0.05$ significance level (these two items were also the items that provide information below the average).

The IRT analyzes were repeated with the remaining eight items after dropping five items that had provided below average information. When the item fit statistics were examined again, it was found that the fit values of the first item were not at an acceptable level of significance ($p<0.05$), and this item was also removed from the scale. As a result of the IRT analysis carried out with the remaining seven items, the item elimination process was terminated by acknowledging that all of the remaining items fit well to the model.

As a result of the primary EFA and CFA analyzes performed, no evidence was obtained regarding the fact that the one-dimensional factor structure of the cyberbullying scale developed by Tanrıkulu, Kınay, and Arıcak (2013) was maintained. After the item removal process, it was confirmed by CFA analysis that the seven-item version showed a one-dimensional structure. In addition, the reliability analysis conducted based on both the classical approach and the IRT showed that the reliability of the scale was similar to the 13-item version. In other words, as a result of the IRT-based item elimination, the construct validity of the scale approached the desired structure while the reliability level of the scale stayed almost the same.

This study contributed to the literature and experts working in the field of cyber psychology in many different ways. Firstly, as emphasized by different experts, understanding the importance of the concept of cyberbullying sensibility and its relationship to other psychological structures is vital. Even though there are many studies forming the nomological network of cyber victimization and cyberbullying with other psychological constructs (Ang & Goh, 2010; Ojedokun & Idemudia, 2013; Kokkinos, Antoniadou, & Markos, 2014;), there is an absence of literature on the construct of cyber sensibility. In addition, there are appropriate instruments for measuring the cyberbullying sensibility among high school students (Tanrıkulu, Kınay, & Arıcak, 2013) while there is no instrument for such studies to be carried out with university students. This validated instrument will contribute to a better understanding of the cyberbullying sensibility of university students, and our understanding of the relationship between cyberbullying sensibility and other structures will expand for this population.

Secondly, reducing the number of items increased the usefulness of the scale while the amount of information it provides was mostly retained. In addition, the reliability level of the scale was kept at almost similar levels despite item removal. We expect that a shorter but still-reliable scale will be preferred by the researchers and practitioners. On the other hand, the research has some limitations that should be taken into consideration by the readers and scientists who will use this instrument in their research and practices. The data obtained in this study were collected only from the students who were studying at a private university. For this reason, there is a question about the generalizability of the findings. The university where the data collection process took place is located in a medium-sized metropolis. Inclusion of university students in larger metropolitan cities such as Istanbul into the data collection process and the inclusion of students in state owned universities will increase the generalizability of the findings. On the other hand, as Baker (2001) states, IRT parameters are independent of the sample data collected. Therefore, the parameters obtained are independent of the sample group. Hence, we think that generalizability of the results may not pose a serious problem based on the fact that the IRT was employed.

Secondly, evidence was obtained for the validation of the cyberbullying sensibility scale for university students. On the other hand, no data collection process was carried out to gather evidence of criterion validity. In addition, no test re-test process was carried out to determine whether the scale gives stable results. We recommend that future studies should focus on collecting evidence for criterion related validity and the stability of scores across time for the 7-item cyberbullying sensibility scale.

Third, the resulting difficulty parameters show that the scale can perform more accurate measurements for the average-to-high level of sensibility. This finding was observed for both the 13-item version and the seven-item version (see Table 1 and 2). When it is considered that the individuals who need preventive interventions by experts have low levels of cyberbullying sensibility, it is seen that it is necessary to add more questions to the scale that can give information for the lower level of cyberbullying sensibility. In the further revisions of the scale, we recommend that authors add items suitable for providing information on individuals with lower cyberbullying sensibility level for risky populations.

Differential Item Functioning (DIF) was not investigated in the current study. DIF occurs when different subgroups of participants (e.g., male and female) with the same latent trait level yield different response patterns. If DIF is detected, it poses a risk to the validity of the scale. Because CSS is a relatively new construct, there is little knowledge about whether some subgroups present more cyberbullying sensibility. Hence, we recommend the investigation of DIF across different subgroups of cyberbullying sensibility. Such an investigation might reveal possible differences in subgroups.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

ORCID

Osman Tolca Arıcak https://orcid.org/0000-0001-8598-5539
Akif Avcu https://orcid.org/0000-0003-1977-7592
Feyza Topçu https://orcid.org/0000-0002-5853-2670
Merve Gülçin Tutlu https://orcid.org/0000-0003-4225-7982

5. REFERENCES

Álvarez-García, D., Núñez, J.C., González-Castro, P., Rodríguez, C., and Cerezo, R. (2019) The Effect of Parental Control on Cyber-Victimization in Adolescence: The Mediating Role of Impulsivity and High-Risk Behaviors. *Front. Psychol*., 10, 1159.

Ang, R.P., & Goh, D.H. (2010). Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender. *Child Psychiatry & Human Development*, *41*(4), 387-397.

Bilker, W.B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354-369.

Baker, F.B. (2001). *The basics of item response theory (2nd ed.)*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland. Retrieved February, 3 2019 from http://files.eric.ed.gov/fulltext/ED458219.pdf

Baştak, G., & Altınova, H.H. (2015). Lise Öğrencilerinde Yaratıcı Drama Yöntemiyle Siber Zorbalık Hakkında Duyarlılık Oluşturma. *Yaratıcı Drama Dergisi*, *10*(1), 91-102.

Chalmers, R.P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1-29.

Doğan, E., Cansu, Ç., & Şahin, Y.L. (2016). A Study on Online Social Network Games Players' Cyberbullying Sensibility and Aims of Facebook Usage/Çevrimiçi Sosyal Ağ Oyunu Oynayan Bireylerin Siber Zorbaliğa Duyarlılık Düzeyleri ile Facebook Kullanım Amaçları Üzerine Bir Çalışma. *Eğitimde Kuram ve Uygulama*, *12*(3), 501-520.

Embretson, S., Reise, S.P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates. Inc. Mahwah.

Gaffney, H., Farrington, D. P., Espelage, D. L., and Ttofi, M. M. (2019). Are cyberbullying intervention and prevention programs effective? a systematic and meta-analytical review. *Aggress. Violent Behav. 45*, 134–153. doi: 10.1016/j.avb.2018.07.002

Gahagan, K., Vaterlaus, J.M., & Frost, L.R. (2016). College student cyberbullying on social networking sites: Conceptualization. prevalence. and perceived bystander responsibility. *Computers in human behavior*, *55*, 1097-1105.

Hambleton, R.K, Swaminathan, H., Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks: Sage Publications.

Hambleton, R.K., Robin, F., & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In: Tinsley HEA. Brown SD. editors. *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. San Diego: Academic. p. 553–85.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Academic Publishers.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*, 139–64.

Hu. L. T. & Bentler. P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Istiyono, E., Dwandaru, W.S.B., Lede, Y.A., Rahayu, F., & Nadapdap, A. (2019). Developing IRT-Based Physics Critical Thinking Skill Test: A CAT to Answer 21st Century Challenge. *International Journal of Instruction*, *12*(4), 267-280.

Kokkinos, C.M., Antoniadou, N., & Markos, A. (2014). Cyber-bullying: An investigation of the psychological profile of university student participants. *Journal of Applied Developmental Psychology*, *35*(3), 204-214.

Lee, J., Abell, N., & Holmes, J.L. (2015). Validation of measures of cyberbullying perpetration and victimization in emerging adulthood. *Research on Social Work Practice*. http://dx.doi.org/10.1177/1049731515578535

Mielenz, T.J., Edwards, M.C. & Callahan, L.F. (2010). Item response theory analysis of two questionnaire measures of arthritis-related self efficacy beliefs from community based US samples. *Hindawi Publishing Corporation Arthritis*.

Muthén, L.K., & Muthén, B.O. (1998-2012). *Mplus User's Guide: Statistical Analysis with Latent Variables* (7th ed.). Los Angeles. CA: Muthén & Muthén.

Nedim-Bal, P., & Kahraman, S. (2015). The Effect of Cyber Bullying Sensibility Improvement Group Training Program on Gifted Students. *Journal of Gifted Education Research*, *3*(2). 48-57.

Ojedokun, O., & Idemudia, E. S. (2013). The moderating role of emotional intelligence between PEN personality factors and cyberbullying in a student population. *Life Science Journal*, *10*(3), 1924-1930.

Orlando, M., & Thissen, D. (2000). New item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*, 50-64.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Austria. URL https://www.R-project.org/

Rubio, V.J., Aguado, D., Hontangas, P.M., & Hernandez, J.M. (2007). Psychometric properties of an emotional adjustment measure. *European Journal of Psychological Assessment*, 23 (1), 39-46.

Samejima, F. (1969) Estimation of Latent Ability Using a Response Pattern of Graded Scores. (Psychometrika Monograph, No. 17). Psychometric Society, Richmond. http://www.psychometrika.org/journal/online/MN17.pdf

Tanrıkulu, I. (2018). Cyberbullying prevention and intervention programs in schools: A systematic review. *School psychology international*, *39*(1), 74-91.

Tanrıkulu, T., Kınay, H. & Arıcak, O.T. (2013). Cyberbullying sensibility scale: validity and reliability study. *Trakya University Journal of Education*, 3(1), 38-47.

Tanrıkulu, T., Kınay, H., & Arıcak. O. T. (2015). Sensibility development program against cyberbullying. *New Media & Society*, 17(5), 708-719.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51(4), 567–577.

Tolia, A. (2016). Cyberbullying: Psychological effect on children. *The International Journal of Indian Psychology*, *3*(2), No. 1. 48-51.

Tsutakawa, R.K., & Johnson. J.C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, *55*, 371–390.

Willoughby, M. (2018). A review of the risks associated with children and young people's social media use and the implications for social work practice. *Journal of Social Work Practice*. 1-14. doi:10.1080/02650533.2018.1460587

Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three parameter logistic model. *Applied Psychological Measurement*, *8*, 125-145.

Zanon, C., Hutz, C.S., Yoo, H., & Hambleton, R.K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, *29*, 1-10.