Asian Nursing Research 13 (2019) 249–256

Contents lists available at ScienceDirect

Asian Nursing Research

journal homepage: www.asian-nursingresearch.com

Research Article

An Item Response Theory Analysis of the Korean Version of the CRAFFT Scale for Alcohol Use Among Adolescents in Korea

Youngshin Song,[1] Hyerang Kim,[2] So-Youn Park[3,*]

[1] College of Nursing, Chugnam National University, Daejeon, Republic of Korea
[2] Red Cross College of Nursing, Chung-Ang University, Seoul, Republic of Korea
[3] Division of Human Service, Kyonggi University, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Received 12 April 2019
Received in revised form
11 September 2019
Accepted 23 September 2019

Keywords:
adolescent
alcohols
psychometrics
validation studies

ABSTRACT

Purpose: This study aimed to validate the psychometric properties of the CRAFFT (Car, Relax, Alone, Forget, Family/Friends, Trouble) by using item response theory (IRT) and further examine gender differences in item-level responses.
Methods: This study used the 13th (2017) Korea Youth Risk Behavior Survey data conducted by the Korean Centers for Disease and Prevention and analyzed data of 8,568 students who reported drinking alcohol in the previous 30 days. IRT assumptions including unidimensionality, local independence, and monotonicity were tested. A one-factor two-parameter IRT model was fitted for item parameterization ($M_2 = 76.92$, 20 df, $p < .001$, root mean square error of approximation = 0.02). Graphics for item characteristic curves and item and test information curves were provided. Differential item functioning (DIF) analysis was evaluated to measure item equivalence between boys and girls.
Results: Unidimensionality and local independence were satisfied, and the dominant factor eigenvalue and variance were 2.48 and 41.3% in boys and 2.08 and 34.7% in girls, respectively, and residual correlations in paired items were less than .2. The discrimination parameters were estimated as 1.32–3.59 for boys and 1.30–3.64 for girls. The severity parameters estimated verity as −0.23 to 1.67 for boys and −0.32 to 2.17 for girls. DIF was detected for four of six items, "Car," "Relax," "Forget," and "Family/friends." However, its impact on the total score was negligible.
Conclusion: The CRAFFT is a valid and reliable screening tool to identify alcohol use problems for both boys and girls, exhibiting good discrimination, good coverage of severity, and negligible DIF.

Introduction

Adolescent drinking is a common social issue in all countries. Alcohol use in adolescents is associated with serious negative consequences, such as self-harm, suicidal attempts, truancy, delinquency, unintended pregnancy, and academic failure [1,2]. Furthermore, adolescents who drink at an early age are at higher risk of developing alcohol use problems and disorders [3].

The 2018 Korea Youth Risk Behavior Survey (KYRBS) reported that 18.7% of boys and 14.9% of girls in Korea aged between 13 to 18 currently drink, among whom 37.2% of the boys and 38.9% of the girls showed problematic drinking behaviors [4]. These numbers for adolescent drinkers in Korea are higher than those in other countries as the United Kingdom, in which the adolescent drinking rate was 9.9%, as found in the Millennium Cohort Study [1]. Such statistics suggest that Korean drinking adolescents might face even higher risks for alcohol abuse.

Although the Korean government has been enforcing a strict policy prohibiting supplying alcohol to adolescents and conducting rigorous public education and campaigns against alcohol abuse among adolescents [5], adolescent alcohol consumption is not decreasing [5]. These circumstances have led practitioners and clinicians to pay more attention to the importance of early intervention and treatment, such as early identification of adolescent alcohol abuse.

Youngshin Song: https://orcid.org/0000-0002-2203-5840; Hyerang Kim: https://orcid.org/0000-0002-8044-2188; So-Youn Park: https://orcid.org/0000-0001-8386-6245

* Correspondence to: So-Youn Park, Ph.D., Division of Human Service, Kyonggi University, San 94-6, Iui-dong, Yeongtong-gu, Suwon-si, Kyonggi-do, 15442, Republic of Korea.
E-mail address: spark831@gmail.com

https://doi.org/10.1016/j.anr.2019.09.003

The Korea Centers for Disease Control and Prevention (KCDC) has conducted the KYRBS annually since 2005 to examine the health status of Korean middle- and high-school students [4]. The KYRBS surveys 16 areas of health and risk behaviors, including mental, physical, oral, and reproductive health; substance abuse; internet addiction; dietary and bodyweight control; and personal hygiene [4]. The KYRBS is conducted nationally, and its data enable health indicators to plan and evaluate youth health–related projects sponsored by the Korean government.

To monitor problematic drinking behaviors among Korean adolescents, the KYRBS uses the CRAFFT (an acronym for the first letter of the key words Car, Relax, Alone, Forget, Family/Friends, Trouble [6]). The CRAFFT, which is a six-item instrument, was developed by Knight et al. [6] to screen problematic alcohol use in adolescents. Although there are many other instruments to diagnose alcohol use problems, such as the Alcohol Use Disorders Identification Test [6], the Problem Oriented Screening Instrument for Teenagers [7], and the CAGE (an acronym for the first letter of key words in Cut down, Annoyed, Guilty, Eye Opener) [8], the CRAFFT has many advantages over the other tools. First, the CRAFFT is specifically designed for adolescents [9]. Second, the CRAFFT is time-efficient because testing takes approximately three minutes to complete, which makes it easy to administer as an interview or self-report [10].

Currently, the CRAFFT has been translated into different languages. To serve as a valid screening tool for alcohol use problems in adolescents in different parts of the world, some adjustments have been necessary. For example, the Argentine [11], French [12], Norwegian [13], Turkish [14], and Korean versions of the CRAFFT [15] have been evaluated for sensitivity and specificity, showing that there were variations in sensitivity and specificity, which suggests that these results were affected by each country's own societal and cultural factors. As for the Korean version of the CRAFFT, the question related to “Car” (“have you ever ridden in a car driven by someone including yourself who was high or had been using alcohol or drugs?”) was not sensitive enough to identify problematic alcohol use in Korean adolescents, suggesting that this item could be modified [15]. Similarly, in the Argentinian version of the CRAFFT, questions related to “Car” and “Alone” (e.g., “Do you ever use alcohol or drugs while you are by yourself, or alone?”) had little influence on screening for alcohol use problems [11]. Adaptation of the CRAFFT to each country's unique circumstances is necessary to ensure that the measure is culturally and developmentally appropriate for its adolescent population [14].

These aforementioned studies that validated different language versions of the CRAFFT used classical test theory (CTT), which involves examining reliability and factor structure [11,12,14,15], whereas there is a lack of studies conducted to validate the CRAFFT using item response theory (IRT) [13]. To validate an instrument, IRT has some advantages in overcoming the limitations of CTT, for which reason this study used IRT. The following section includes two parts: (1) an overview of IRT and (2) the aims of this study.

Overview of IRT

In general, a measure is composed of multiple items; IRT focuses on the item level of a measure, while CTT focuses on the full scale [16]. IRT can evaluate how each item functions within a scale and how much information each item contributes to a scale [17]. IRT, which developed from CTT, has some advantages over CTT. First, IRT computes item statistics that are independent of the sample from which they were estimated [18]. Second, a respondent's scores in a measure are independent of item difficulty [18], that is, independent of the particular choice of item. Third, in IRT, the standard error of measurement is not assumed to have a constant value regardless of the respondent's location on the range of a scale [18]. This means that a measure is not equally precise across the full range of possible item scores [19]. For example, respondents' highest scores on a measure will not have the same level of precision as their scores in the middle of the measure [20].

IRT uses a statistical model to provide item characteristics, including item difficulty, item discrimination, and differential item functioning (DIF) [21]. Item difficulty is an arbitrary value of proficiencies or traits that measures how difficult the item is [22]. For example, when the response format of a scale is dichotomous, item difficulty identifies the proficiency at which about 50% of respondents are expected to answer the item correctly. In addition, item discrimination identifies how well an item can differentiate between respondents with different trait levels [19]. Higher discrimination means that the item differentiates well between respondents with different levels of the trait measured by it. Moreover, IRT provides DIF, that is, an identification of items for bias that evaluates true differences for different populations (e.g., gender or ethnicity) [22]. One of IRT's assumptions is that item parameters are consistent among subgroups. Thus, DIF is a powerful IRT statistic to help identify biased items.

To sum up, IRT can provide a critical view of item functioning without sample and test biases and the impact of item biases across different populations [19,22]. Applying IRT to analyze the CRAFFT scale can give us greater insight into the functioning of each item as well as the full scale and the different meanings that an item may have for different subgroups.

Aim of the study

The CRAFFT was originally developed to screen for alcohol and drug use problems. According to the 2017 KYRBS, the prevalence of current drinking in adolescents was 16.9%, whereas that of current drug use was 1.0% [23]. These statistics show that adolescents have more problems with alcohol than drug use, and a possible reason for the higher use of alcohol use than drugs is that adolescents have easier access to alcohol than drugs. Accordingly, as alcohol is more readily available, the government pays more attention to alcohol use problems. Given these circumstances, the CRAFFT questionnaires in the KYRBS ask only about alcohol use, and the focus of the present study is on alcohol use problems in the CRAFFT.

To our knowledge, only one study has used IRT to evaluate the CRAFFT in a representative Norwegian adolescent population [13]; however, this study did not provide details of item response and functioning, presenting only difficulties across items. In addition, most previous validation studies have been conducted in hospital settings [9–11] or with clinical patient samples [12], yet little research has been conducted in a school-based adolescent population [13]. Furthermore, substantial evidence shows variations in alcohol use patterns, problems, and motives by gender [24–26], yet a paucity of research is available examining gender differences in item-level responses in the CRAFFT. It is important that any screening instrument functions well regardless of setting and gender [27].

Gender differences in drinking patterns and drinking motives have been found in several studies [25,26]. Male adolescents reported drinking for social or enjoyment enhancement reasons, while female adolescents tend to drink as a coping mechanism to deal with their emotions [25]. Thus, it is necessary to examine whether CRAFFT indiscriminately identifies drinking problems in both boys and girls.

Overall, the aims of this study were twofold. First, this study examined the psychometrics of the CRAFFT among the general population of Korean adolescents using IRT. Second, this study investigated item equivalence between genders.

Methods

Study design and data source

The study evaluated the psychometric properties of the CRAFFT using the 13th (year 2017) KYRBS data. The KYRBS, which was designed to provide health-risk behavior epidemiological information on middle- and high-school students nationwide, has been conducted annually by the KCDC in Korea since 2005 [28]. The KYRBS collected information during the month of June 2016 from a nationally representative sample of students who attended middle and high schools in 17 provinces.

The KYRBS used a multi-stage cluster probability sampling design to recruit participants. First, the primary sampling units (PSUs) were schools. Second, once PSUs were selected, a systematic sampling method was used to sample one classroom of each grade [28]. All students in the sampled classes were eligible to participate and use a Web-based survey (p. 1076) [28]. The mean time to complete the survey was 45–50 minutes. A more detailed description of the KYRBS data is found on the KCDC website [23].

Participants

The total sample size of the KYRBS was 62,276 adolescents aged 12–18 years. Of the total sample, this study used 8,568 students who reported that they had drunk alcohol in the previous 30 days. The mean age of the participants was 15.94 ± 1.50 (mean ± standard deviation). The participants comprised 4,997 boys (58.3%) and 3,571 (41.7%) girls, and there were 2,070 middle-school students (24.2%) and 6,498 high-school students (75.8%). Approximately 21% answered that they had used alcohol for the first time during elementary school. About 40% answered that they had consumed 5–6 drinks at a time during the last month. There were differences in alcohol consumption between boys and girls: Boys were more likely to binge-drink (consume 5–6 drinks at a time [29]) than girls (46.1% in boys and 37.2% in girls, $p < .01$), and the prevalence of heavy alcohol use (defined as binge-drinking on five or more days in the past month) in boys was higher than in girls (14.4% in boys and 8.5% of girls, $p < .05$).

Measures

The CRAFFT comprises six items that ask about problems related to alcohol use to detect alcohol use problems over the past year [6]. As mentioned previously, the KYRBS modified the CRAFFT questionnaire for the current Korean situation; thus, only alcohol use was asked in the original CRAFFT instrument. As mentioned earlier, the name CRAFFT was formed from the first letters of the key words in six questions: (1) Have you ever ridden in a “Car” driven by someone (including self) who was high or had been using alcohol? (2) Do you ever use alcohol to “Relax,” feel better about yourself, or fit in? (3) Do you ever use alcohol while you are by yourself, “Alone”? (4) Do you ever “Forget” things you did while using alcohol? (5) Do your “Family or friends” ever tell you that you should cut down on your drinking? and (6) Have you ever gotten into “Trouble” with the law while you are using alcohol? The instrument uses dichotomous responses of *yes* (1) or *no* (0), and the sum of scores ranges from 0 to 6. Higher scores indicate greater severity of alcohol use problems.

Ethical considerations

This study involved the secondary analysis of a preexisting, deidentified, and publicly available data set, KYRBS. This study was approved by Chungnam National University (Approval no. 201903-SB-021-01).

Data analysis

Before IRT analysis, a total of 8,568 students who completed the CRAFFT questionnaire were selected for a separated data set. Factor analysis was conducted using eigenvalue and variance. Model fit was tested by examining the difference in -2 log likelihood of the one-parameter and two-parameter model. The difference in -2 log likelihood of two models is distributed as $\chi 2$ values by degrees of freedom (df). A one-factor two-parameter logistic model was used to examine item responses and their relationships with the trait measured by the scale. The essential prerequisites for an IRT model, such as unidimensionality, local independence, and monotonicity, were assessed by examining eigenvalues on a scree plot, residual correlations, and the shape of the item characteristic curve. Factor loadings (λ), discrimination (a) parameters, and difficulty (severity of the alcohol use problem in this study) (b) parameters were estimated and presented with item characteristic curves. The (b) parameter indicated the position of the item characteristic curve on the latent trait continuum. The (a) parameter measured the degree of precision with which an item was distinguished between low-trait and high-trait individuals (lower alcohol use problems vs. higher alcohol use problems in this study). An item and test information curve described the amount of information provided by each item and the items as a set of the test along the severity continuum. The more information a test provides at a particular trait level, the smaller the standard errors in that range. Model fit was determined by S-χ^2 fit statistics, which express the degree of fit or misfit between observed and expected values, along with M_2 fit statistics and root mean square error of approximation in the data.

DIF analysis was conducted to evaluate item variance between genders by examining item parameters that were significantly different between boys (reference group) and girls (test group). An iterative process for DIF analysis was conducted: (1) identifying items without DIF, (2) iterative testing of each item for DIF, and (3) fitting a final model. For the first step, anchor items (items without DIF) and test items (items with DIF) were identified by comparing a parsimonious model, in which all parameters constrained to be equal were set free to be estimated distinctly for the two groups. The Wald statistic was used to test the significance of DIF, yielding a χ^2 value and df for each parameter estimate between the two groups. For each item, the significance of the χ^2 value was considered an indication of DIF. After identifying anchor items, the process was repeated until no test items were identified as having DIF. The marginal maximum likelihoodestimation method of Bock and Aikin was used to fit the final models and estimate the parameters.

The required sample size for conducting IRT analysis including DIF detection is greater than 200 cases [30]. The aforementioned sample size for the present study was 8,568, which was sufficient for analysis. Descriptive statistics analyses of the participants were performed using SPSS version 18 (IBM Corp., Armonk, NY, USA), and other statistical analyses were conducted using IRTPRO version 4.2 (Socientifit Software International, Inc., Lincolnwood, IL, USA).

Results

Model fit

A one-factor two-parameter logistic model was used to examine item responses and their relationships with the trait measured by the scale. The one-factor model was determined with the first eigenvalue (2.48 in boys and 2.08 in girls) and variance (41.3% in boys and 34.7% in girls). Internal consistency was determined by Kuder–Richardson Formula 20 to be .69 and .60, respectively. The difference between -2 log likelihood for the one-parameter and two-parameter models was not significant (40592.30 − 40569.69 = 22.61, 6 df, $p = .070$), which indicates that there is no statistical significance in the difference in discrimination for the items. However, we decided to use a two-parameter model to find more information about the characteristics of items. It was intended to identify respondents' trait for alcohol use to pursue our study aim, which is to evaluate different item function between boys and girls as mentioned earlier.

Table 1 presents the parameters for each item and the χ^2 values and probabilities for the tests of *a* and *b* DIF by gender. The *p*-values for the S-χ^2 fit statistics of the items were all less than .01, and the model had adequate fit ($M_2 = 76.92$, 20 df, $p < .001$, root mean square error of approximation = .02).

Item discrimination

Item discrimination (*a*) parameters indicate the differences in item sensitivity to variations in alcohol use, demonstrating that changes in certain items with larger discrimination may be more clinically meaningful than others. Items on the CRAFFT scale appeared to perform similarly across boys and girls as groups (Figure 1). All items exhibited good discrimination (1.32–3.59 for boys and 1.30–3.64 for girls). "Trouble" had the greatest discrimination ($a = 3.59$ for boys and $a = 3.64$ for girls), indicating that this item was most sensitive in distinguishing individual differences in the severity of alcohol use problems, while "Alone" showed the least discrimination ($a = 1.32$ for boys and $a = 1.30$ for girls) and thus the least sensitivity to individual differences in the severity of alcohol use problems.

Item difficulty

The item difficulty (*b*) parameters indicated differences in levels of severity of alcohol use problems, such that higher *b* estimates on the continuum of θ reflect more severe alcohol use problems. The *b* estimates indicated good coverage of severity and ranged from −0.23 to 1.67 in boys and from −0.32 to 2.17 in girls. In boys, the items with highest scores were "Car" ($b = 1.67$), "Friends" ($b = 1.52$), and "Trouble" ($b = 1.49$), and the lowest score was on "Relax" ($b = -0.23$), which reflected that items with higher *b* estimates may be more clinically relevant to problematic alcohol use. Similarly, in girls, the highest was "Car" ($b = 2.17$) and the lowest was "Relax" ($b = -0.32$), reflecting that "Car" may be more clinically relevant to problematic alcohol use than "Relax."

Differential Item Functioning

Regarding measurement equivalence between boys and girls, four items showed significant DIF: "Relax" exhibited both *a* DIF and *b* DIF, while "Car," "Forget," and "Family/friends" showed only *b* DIF (Table 1). The trace lines for the items "Car," "Relax," "Forget," and "Family/friends" were not identical between boys and girls, presenting left-shifted lines for "Relax," "Forget," and "Family/friends" and a right-shifted line for "Car" in girls compared with boys (Figure 1). These shifts indicate differences between boys and girls in responding "yes" to the items. For example, girls who were less problematic were more likely to respond "yes" for "Relax," "Forget," and "Family/friends" than boys with a similar trait level. There was a reverse trend for "Car": Girls who were more problematic were more likely to respond "yes" to the item than boys with a similar trait level. "Relax" had both *a* DIF and *b* DIF, which suggested that it may sensitively distinguish severity of alcohol use problems among girls with less severe alcohol use problems.

Item and test information functions

Item information curves showed that the information was concentrated around the area defined by the severity parameters (Figure 2). In both groups, "Trouble" was considered the most informative item to identify problematic behavior, especially for adolescents with more problematic behavior, while "Alone" provided the least information. Otherwise, the curve for "Relax" had a reverse trend compared with the other items, functioning relatively well at lower levels of problematic alcohol use, which implies good measurement properties for moderate or lower levels of the continuum. It distinctively contributed more information about less problematic behavior. The test information curve with standard errors of measurement indicates the amount of information provided by the instrument across the continuum of problematic use (Figure 3). The six items of the test performed better as a set at measuring a slightly narrow range of moderately severe levels on the continuum. It also showed that more information was provided for girls at middle-lower levels (e.g., $-1 < \theta < 1$) than for boys at the same level. However, the expected score did not substantially differ between genders across the range of severity of alcohol use problems (Figure 3), indicating that differences in DIF did not essentially change the total score level and canceled out at the test level.

Discussion

This study evaluated the psychometric properties of the CRAFFT using IRT, suggesting that this screening instrument is valid for identifying alcohol problems among Korean adolescents. Given the

Table 1 *Differential Item Functioning by Gender (N = 8,568).*

Items	Boys (n = 4,997)				Girls (n = 3,571)				χ^2 (*p*)	
	Response to yes (%)	λ	*a* (SE)	*b* (SE)	Response to yes (%)	λ	*a* (SE)	*b* (SE)	*a* DIF	*b* DIF
Car	472 (9.4)***	0.79	2.22 (0.13)	1.67 (0.05)	132 (3.7)***	0.86	1.95 (0.25)	2.17 (0.15)	1.0 (.319)	28.9 (<.001)
Relax	2,815 (56.3)***	0.68	1.58 (0.09)	−0.23 (0.03)	2,246 (62.9)***	0.96	2.39 (0.23)	−0.32 (0.13)	11.3 (<.001)	10.1 (.002)
Alone	1,354 (27.1)	0.61	1.32 (0.07)	0.99 (0.05)	913 (25.6)	0.66	1.30 (0.07)	1.00 (0.05)		
Forget	932 (18.7)***	0.79	2.15 (0.12)	1.14 (0.04)	790 (22.1)***	1.04	2.89 (0.44)	0.77 (0.05)	2.6 (.105)	32.9 (<.001)
Family/Friends	657 (13.1)	0.74	1.88 (0.10)	1.52 (0.05)	491 (13.7)	0.95	2.35 (0.27)	1.18 (0.06)	2.7 (.104)	15.7 (<.001)
Trouble	446 (8.9)***	0.91	3.59 (0.34)	1.49 (0.03)	189 (5.3)***	1.12	3.64 (0.34)	1.49 (0.03)		

Note. DIF = differential item functioning; SE = standard error.
***$p < .001$

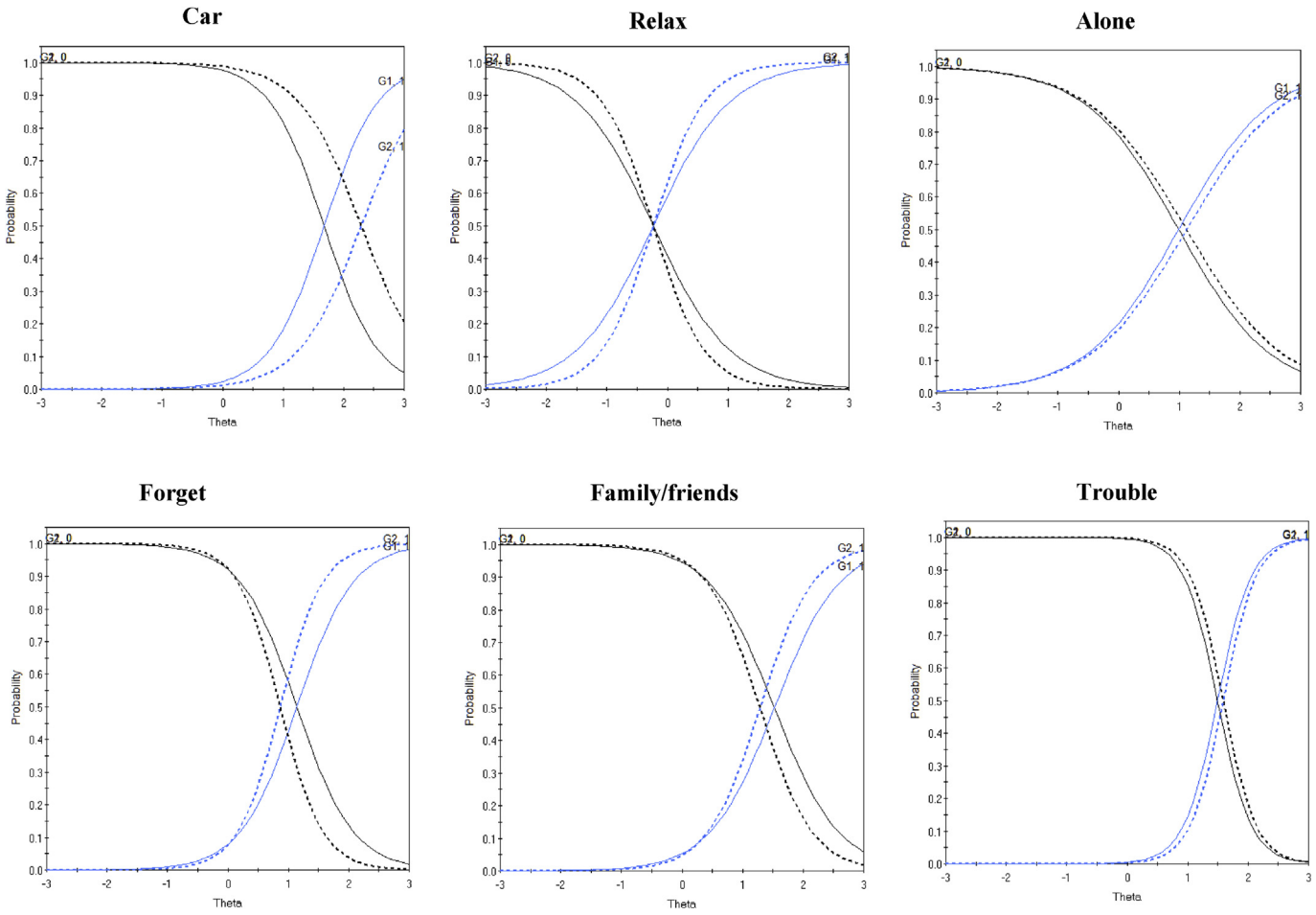

Figure 1. Item characteristic curves for all 6 items.
Note. Solid line (G1) for boys and dotted line (G2) for girls.

brevity of the instrument, CRAFFT appears to be a practical and effective tool to screen for alcohol use problems.

As for the details of the answer responses in this study, the frequency of positive answers regarding alcohol use problems ranged from 8.9% to 56.3% for boys and from 3.7% to 62.9% for girls. Of the six items, “Car” (“Have you ever ridden in a car driven by someone including yourself who was high or had been using alcohol?”) had the least positive responses by both groups (9.4% and 3.7% for boys and girls, respectively). However, other studies showed a higher percentage of positive responses to the “Car” question, such as 68% in the study by Bernard et al. [12] conducted in Switzerland, and 42.6% in the study by Knight et al. [9] conducted in the U.S. The differences in findings may be due to different sociocultural factors, such as differences in the legal ages for drinking and driving. The legal drinking age in Korea is 19 years, while in Switzerland, teenagers can legally drink alcohol at the age of 15 years and at 21 years in the U.S. [12,15]. At the same time, a driver's license is issued to persons aged 19 years and older in Korea, while in Switzerland and in the U.S., the legal driving age is 18 years and 16 years, respectively [12,15]. The combination of different legal drinking and driving ages serves as a sociocultural variable. The Korean legal drinking and driving ages are higher than in the other two countries, allowing less access to alcohol and occasions these legal-aged Korean teenagers to drive in their younger years, which might have contributed to lower positive responses to “Car.”

In contrast, “Relax” (“Do you ever use alcohol to relax, feel better about yourself, or fit it?”) received the highest rate of positive responses by both groups (56.3% and 62.9% for boys and girls, respectively). For this item, Bernard et al. [12] and Knight et al. [7] found positive response rates of 73.0% and 15.6%, respectively. These mixed findings may indicate differences in drinking attitudes and patterns influenced by each country's distinct alcohol policy and drinking culture, implying that CRAFFT needs to incorporate an age-specific and culturally appropriate context.

IRT is a powerful technique to test items for bias using DIF analysis, and in this study, gender-related item response bias was examined with DIF. Gender differences in item response were found for “Car,” “Relax,” “Forget,” and “Family/Friend.” This finding may be explained by gender differences in drinking patterns and drinking motives. Several studies have shown that boys were more likely to drink frequently and be drunk than girls [25,26] and that boys tend to drink for enjoyment and for social reasons [25], whereas girls tend to use alcohol to cope with negative emotions, such as depression or anxiety, but are less likely to drink in response to peer pressure or to fit into a group [31]. The findings of our DIF analysis suggest that boys and girls have different drinking patterns and motives. Regarding the “Relax” question which tried to track down the motivation for drinking, girls responded that they drink for relaxation more than boys do. Thus, differences in motives and attitudes toward alcohol consumption might have yielded a different response to this item in CRAFFT. It also appears that the discriminative power of this item was higher in girls than in boys. Regardless of the apparent gender differences in the item properties, DIF was not reflected at the test level, meaning that DIF did not

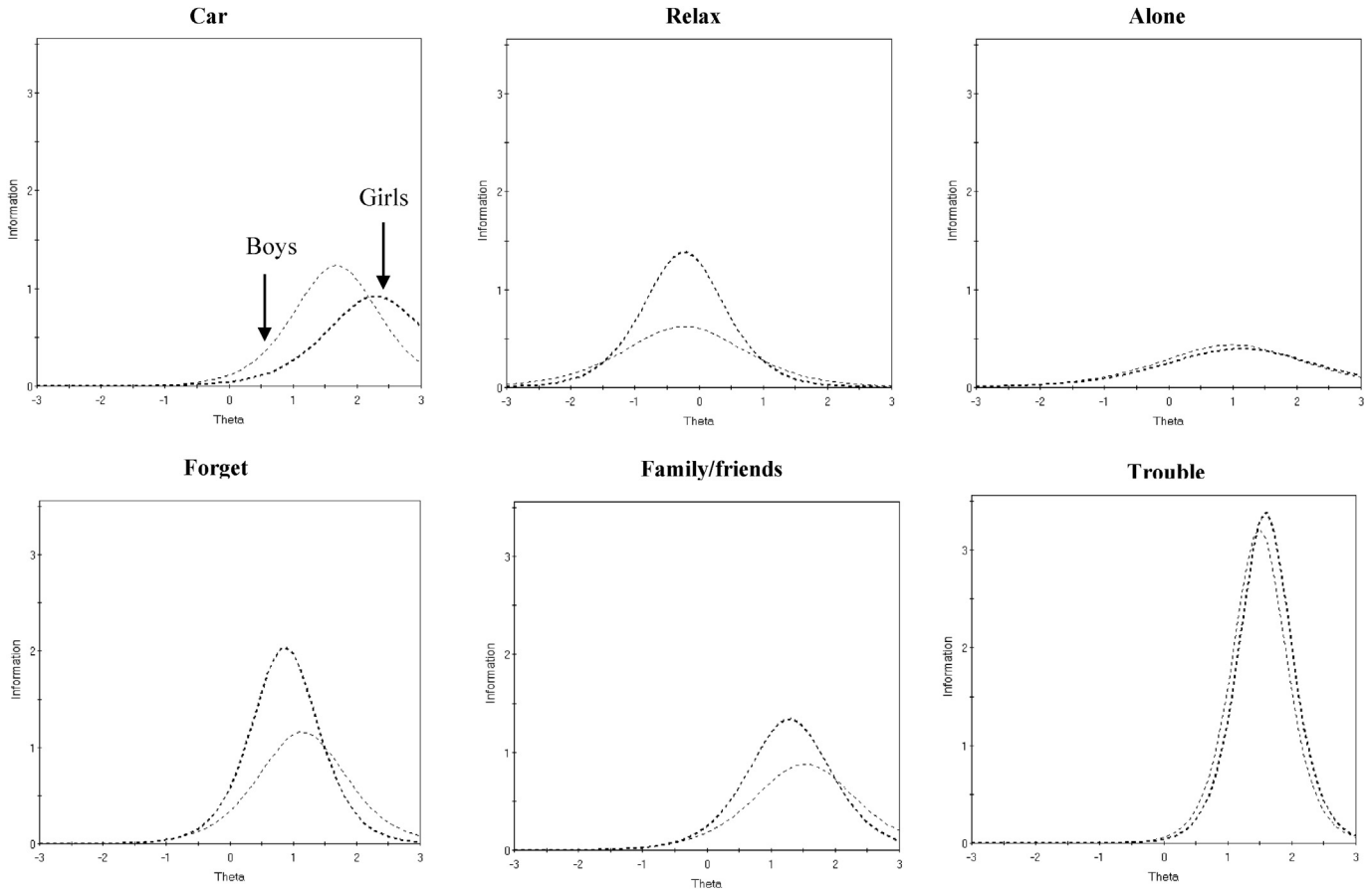

Figure 2. Item information curves for all 6 items in the CRAFFT scale.
Note. Lightface dotted line for boys and boldface dotted line for girls.

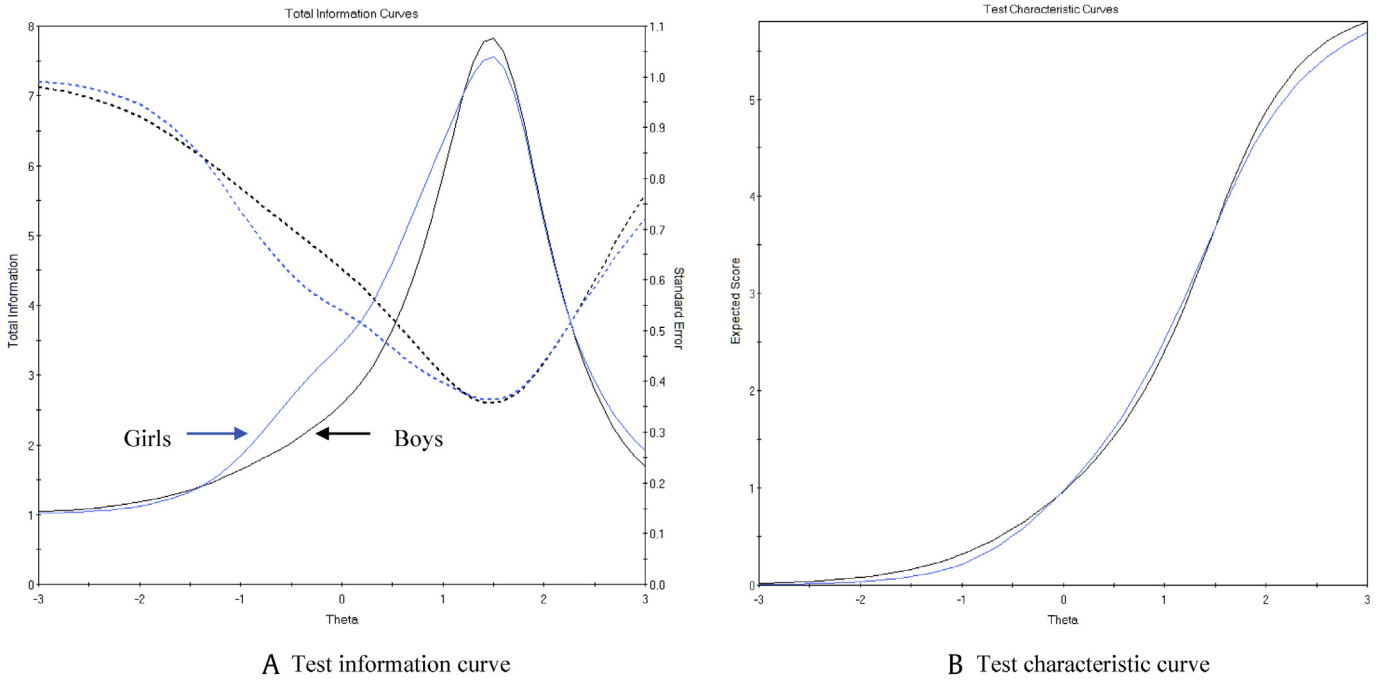

Figure 3. Test information and characteristic curves.
Note. Black line for boys and gray line for girls; solid line for total information and dotted line for standard error.

affect gender differences overall. In other words, as the six items of the CRAFFT, whose scores were summed to yield the total score for the whole scale, did not yield different results for boys and girls, the CRAFFT is confirmed to be suitable as a screening tool to identify at-risk alcohol users among Korean youth.

Several limitations of this study should be noted. First, there is a limitation in the measurement that might have biased the results. Specifically, the "Relax" question ("Do you ever use alcohol to Relax, feel better about yourself, or fit it?") may be taken as asking about two different drinking attitudes at the same time, that is, drinking for enjoyment and for social reasons [12]. It has been suggested that these two different drinking attitudes should be asked about separately. Second, owing to its nature as a secondary analysis, the present study was limited by the sample. The KYRBS did not include adolescents who were not enrolled in school (e.g., drop-outs). Thus, the findings of this study apply only to adolescents who attend school, not to those who have left. Third, owing to the nature of self-reporting to measure alcohol use problems, it is unclear how honestly the participants answered the questions. Some adolescents might have overstated their alcohol use, while others might have understated it, which might have biased the findings. Finally, although this study used data with a complex sampling design, we only used the adolescents who completed the CRAFFT in a separate data set without multilevel structure or sample weights. Thus, it must be kept in mind that the study results have a potential bias in the estimates. Because the impact of test length, number of examinees, and values of item parameters vary widely from one context to another [32], it has been recommended that IRT analyses with a complex sample design take a hybrid aggregated–disaggregated approach, which could address biased characteristics of the items and examinees [33]. However, it is known that hybrid methods have a limited selection of specific statistical software (e.g., Mplus) among current standard IRT packages, and there is a lack of empirical analysis applying the method. Further studies performing data analysis with weights are needed.

Despite these study limitations, this study has certain strengths. This study used the KYRBS, a large nationally representative sample of Korean general adolescents, suggesting that the findings can be generalized to the in-school adolescent population. In addition, this study used IRT, including the DIF method, and demonstrated differences in item responses between boys and girls. These findings may be useful for practitioners and clinicians to consider gender-specific responses to drinking patterns and motives in the assessment of alcohol use problems.

Conclusion

The results of this study provide evidence that the CRAFFT is a reliable and practical screening tool to detect alcohol use problems among Korean adolescents, showing high discrimination, good coverage of severity, and negligible item response biases by gender. The findings suggest that certain items function differently in boys and girls, implying that approaches that consider gender differences are needed to reduce alcohol-related risky behavior.

Funding

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea [NRF-2017S1A5B8057479].

Declaration of competing interest

The authors have declared no conflict of interest.

References

1. Patalay P, Gage SH. Changes in millennial adolescent mental health and health-related behaviours over 10 years: a population cohort comparison study. Int J Epidemiol. Forthcoming. 2019. https://doi.org/10.1093/ije/dyz006
2. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics. 2007;119(1):76–85. https://doi.org/10.1542/peds.2006-1517
3. Cummins LH, Chan KK, Burns KM, Blume AW, Larimer M, Marlatt GA. Validity of the CRAFFT in American-Indian and Alaska-Native adolescents: screening for drug and alcohol risk. J Stud Alcohol. 2003;64(5):727–32. https://doi.org/10.15288/jsa.2003.64.727
4. Ministry of Health and Welfare of Korea, Centers for Disease and Prevention, Ministry of Education. The 14th (2018) Korean youth risk behavior survey statistics. Sejong City, Korea: MHWK; 2018. Korean.
5. Kim J, Chung W, Lee S, Park C. Estimating the socioeconomic costs of alcohol drinking among adolescents in Korea. J Prev Med Public Health. 2010;43(4):341–51. https://doi.org/10.3961/jpmph.2010.43.4.341. Korean.
6. Knight JR, Shrier LA, Bravender TD, Farrell M, Vander Bilt J, Shaffer HJ. A new brief screen for adolescent substance abuse. Arch Pediatr Adolesc Med. 1999;153(6):591–6. https://doi.org/10.1001/archpedi.153.6.591
7. Rahdert ER. The adolescent assessment/referral system manual [Internet]. Bethesda, MD & Wachington, DC: Pacific Inst. for Research and Evaluation & Westover Consultants, Inc.; 1991 [cited 2019 Mar 10]. Available from: https://files.eric.ed.gov/fulltext/ED340960.pdf
8. Bush B, Shaw S, Cleary P, Delbanco TL, Aronson MD. Screening for alcohol abuse using the CAGE questionnaire. Am J Med. 1987;82(2):231–5. https://doi.org/10.1016/0002-9343(87)90061-1
9. Knight JR, Sherritt L, Shrier LA, Harris SK, Chang G. Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. Arch Pediatr Adolesc Med. 2002;156(6):607–14. https://doi.org/10.1001/archpedi.156.6.607
10. Knight JR, Sherritt L, Harris SK, Gates EC, Chang G. Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcohol Clin Exp Res. 2003;27(1):67–73. https://doi.org/10.1111/j.1530-0277.2003.tb02723.x
11. Bertini MC, Busaniche J, Baquero F, Eymann A, Krauss M, Paz M, et al. Transcultural adaptation and validation of the CRAFFT as a screening test for problematic alcohol and substance use, abuse and dependence in a group of Argentine adolescents. Arch Argent Pediatr. 2015;113(2):114–8. https://doi.org/10.5546/aap.2015.114
12. Bernard M, Bolognini M, Plancherel B, Chinet L, Laget J, Stephan P, et al. French validity of two substance-use screening tests among adolescents: a comparison of the CRAFFT and DEP-ADO. J Subst Abus. 2005;10(6):385–95. https://doi.org/10.1080/14659890412331333050
13. Skogen JC, Bøe T, Knudsen AK, Hysing M. Psychometric properties and concurrent validity of the CRAFFT among Norwegian adolescents. Ung@hordaland, a population-based study. Addict Behav. 2013;38(10):2500–5. https://doi.org/10.1016/j.addbeh.2013.05.002
14. Kandemir H, Aydemir O, Ekinci S, Selek S, Kandemir SB, Bayazit H. Validity and reliability of the Turkish version of CRAFFT substance abuse screening test among adolescents. Neuropsychiatric Dis Treat. 2015;11:1505–9. https://doi.org/10.2147/ndt.S82232
15. Kim YS. Validation of a Korean version of the CRAFFT for screening substance-abusing adolescents. Ment Health Soc Work. 2010;34:30–55. Korean.
16. Ong CW, Pierce BG, Woods DW, Twohig MP, Levin ME. The acceptance and action questionnaire - II: an item response theory analysis. J Psychopathol Behav Assess. 2019;41(1):123–34. https://doi.org/10.1007/s10862-018-9694-2
17. Wu LT, Pan JJ, Blazer DG, Tai B, Stitzer ML, Brooner RK, et al. An item response theory modeling of alcohol and marijuana dependences: a National Drug Abuse Treatment Clinical Trials Network study. J Stud Alcohol Drugs. 2009;70(3):414–25. https://doi.org/10.15288/jsad.2009.70.414
18. Bichi AA, Talib R. Item response theory: an introduction to latent trait moles to test and item development. Int J Eval Res Educ. 2018;7(2):142–51. https://doi.org/10.11591/ijere.v7i2.12900
19. Harvey RJ, Hammer AL. Item response theory. Counsel Psychol. 1999;27(3):353–83. https://doi.org/10.1177/0011000099273004
20. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value Health. 2015;18(1):25–34. https://doi.org/10.1016/j.jval.2014.10.005
21. Scherbaum CA, Cohen-Charash Y, Kern MJ. Measuring general self-efficacy: a comparison of three measures using item response theory. Educ Psychol Meas. 2006;66(6):1047–63. https://doi.org/10.1177/0013164406288171
22. Nguyen TH, Paasche-Orlow MK, Kim MT, Han HR, Chan KS. Modern measurement approaches to health literacy scale development and refinement: overview, current uses, and next steps. J Health Commun. 2015;20(Suppl 2):112–5. https://doi.org/10.1080/10810730.2015.1073408
23. Korea Centers for Disease Control and Prevention (KCDC). The 13th (2017) Korean youth risk behavior survey data [Internet]. Osong, Korea: KCDC; 2017 [cited 2019 Mar 10]. Available from: https://www.cdc.go.kr/yhs/home.jsp?id=m01_01

24. Jackson KM, Sher KJ, Cooper ML, Wood PK. Adolescent alcohol and tobacco use: onset, persistence and trajectories of use across two samples. Addiction. 2002;97(5):517–31.
25. Kuntsche E, Wicki M, Windlin B, Roberts C, Gabhainn SN, van der Sluijs W, et al. Drinking motivates mediate culture differences but not gender differences in adolescent alcohol use. J Adolesc Health. 2015;56(3):323–9. https://doi.org/10.1016/j.jadohealth.2014.10.267
26. Fuller-Thomson E, Sheridan MP, Sorichetti C, Mehta R. Underage binge drinking adolescents: sociodemographic profile and utilization of family doctors. ISRN Fam Med. 2013;2013:1–9. https://doi.org/10.5402/2013/728730
27. Cook RL, Chung T, Kelly TM, Clark DB. Alcohol screening in young persons attending to a sexually transmitted disease clinic: comparison of AUDIT, CRAFFT, and CAGE instruments. J Gen Intern Med. 2005;20(1):1–6. https://doi.org/10.1111/j.1525-1497.2005.40052.x
28. Kim Y, Choi S, Chun C, Park S, Khang YH, Oh K. Data resource profile: the Korea youth risk behavior web-based survey (KYRBS). Int J Epidemiol. 2016;45(4): 1076–1076e. https://doi.org/10.1093/ije/dyw070
29. Drinking levels defined [Internet]. Bethesda, MD: NIAAA; 2018 [cited 2019 Mar 24]. Available from: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking
30. Orlando M, Marshall GN. Differential item functioning in a Spanish translation of the PTSD Checklist: detection and evaluation of impact. Psychol Assess. 2002;14(1):50–9. https://doi.org/10.1037/1040-3590.14.1.50
31. Kuntsche E, Knibbe R, Gmel G, Engels R. Why do young people drink? A review of drinking motives. Clin Psychol Rev. 2005;25(7):841–61. https://doi.org/10.1016/j.cpr.2005.06.002
32. Bulut O, Sünbül Ö. Monte Carlos Simulation Studies in item response theory with the R programming language. J Meas Eval Educ Psychol. 2017;8(3): 266–87. https://doi.org/10.21031/epod.305821
33. Zhen X, Yang JS. Using sample weights in item response data analysis under complex sample designs. In: van der Ark L, Bolt D, Wang WC, Douglas J, Wiberg M, editors. Quantitative psychology research. Springer proceedings in mathematics & statistics. vol. 167. Cham, Switzerland: Springer; 2016. p. 123–37. https://doi.org/10.1007/978-3-319-38759-8_10