

Psychometric Validation of the Persian Version of the Cannabis Use Disorder Identification Test

Mahmood Amirinia^{1*}, Benjamin Ghasemi², Parisa Aghazadeh³, Alireza Mollazadeh⁴

¹Faculty of Medicine Sciences, West Tehran Branch of Islamic Azad University, Tehran, Iran

²School of Environment and Natural Resources, The Ohio State University, Columbus, Ohio, USA

³Department of Psychology and Social Science, Central Branch of Islamic Azad University, Tehran, Iran

⁴Department of Psychology, Ashtian Branch of Islamic Azad University, Ashtian, Iran

Email: *mahmood.amirinia@gmail.com, *mahmood.amirinia@iau.ir

How to cite this paper: Amirinia, M., Ghasemi, B., Aghazadeh, P. and Mollazadeh, A. (2025) Psychometric Validation of the Persian Version of the Cannabis Use Disorder Identification Test. *Open Journal of Medical Psychology*, 14, 71-96.
<https://doi.org/10.4236/ojmp.2025.141005>

Received: December 19, 2024

Accepted: January 27, 2025

Published: January 30, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The purpose of this paper is to look into how reliable and valid the Persian version of the Cannabis Use Disorder Identification Test-Revised (CUDIT-R-Pr) is. It will also compare the screening features of the CUDIT-R with those of the DSM-5 criteria for cannabis use disorder (CUD) based on the SCID-5-CT in a group of university students in Tehran, Iran. The study used the stratified random sampling technique to collect data from 541 students (19 to 24 years old) who used cannabis in Tehran universities in 2024. Confirmatory factor analysis confirmed the uni-dimensionality of the CUDIT-R-Pr. We checked the reliability of the CUDIT-R-Pr using Cronbach Alpha, split-half, inter-rater, test-retest stability over time, and parallel testing equivalence. The results indicated that CUDIT-R-Pr is reliable, reproducible, and responsive, with substantial agreement and adequate interpretability. The CUDIT-R shows that it can tell the difference between different levels of cannabis use severity, which is known as discriminant validity. Receiver operating characteristic analyses confirmed this, using an area under the receiver operating characteristics curve (AUC = 0.95) at a cutoff of ten or less. This allowed CUDIT-R-Pr to accurately predict any DSM-5 based on the highest correctly classified value (0.89), demonstrating high levels of sensitivity (0.96), specificity (0.69), and Youden value (0.65). The exact maximum Youden index (0.72) showed that CUDIT-R-Pr could also predict moderate DSM-5 with a cutoff of twelve or less. To validate and generalize the CUDIT-R-Pr for use among Iranian cannabis users, we need more research.

Keywords

Cannabis Use Disorder, Marijuana, Psychometrics, Validity, Reliability, Iran

1. Introduction

The growing prevalence of cannabis use worldwide, particularly among adolescents and young people, underscores the importance of using a standard and appropriate screening tool to measure cannabis use disorder (CUD) among cannabis (also known as Hashish or Gol) users [1]. Developed countries such as the United States, Canada, and Europe have realized the importance of having such a standard tool, and the high level of research in this field reflects this significance. Developing countries like Iran do not have accurate statistics on the status of cannabis users. The latest statistics published as a household survey on the prevalence of drug and psychotropic use showed that cannabis use after opium was the most-used substance in Tehran province, the capital of Iran, in 2022 [2]. The latest statistics relate to a systematic study among the 21 eastern Mediterranean countries, which demonstrated that the prevalence of CUD in Iran will increase by 4.9%, especially in young men [3]. Throughout addition to these problems and the absence of a common diagnostic instrument for assessing CUD, a common instrument for measuring and identifying CUD throughout Iran is essential.

In this study, we aimed to test and validate a Persian version of the Cannabis Use Disorders Identification Test. We extensively evaluated the CUDIT-R for its psychometric properties and for identifying cannabis use problems. Schultz (2019) [4] and Adamson (2010) [5] found the CUDIT-R reliable and valid, with good internal consistency and concurrent validity. However, Loflin (2018) [6] and Annaheim (2010) [7] noted that the single-factor structure might not be suitable for all sub-populations and suggested further psychometric work. Marshall (2013) [8] and Bonn-Miller (2016) [9] proposed community-based cut points as well as a shortened version to improve its clinical utility. Risi (2020) [10] further validated the CUDIT-R. Coelho (2024) [11] evaluated the accuracy of the CUDIT-R in distinguishing between young adults with and without cannabis use disorder, finding it to be a valid screening tool with excellent sensitivity and specificity. As a result, the CUDIT-R is a widely used screening tool for identifying CUDs and related problems.

In the present study, we evaluated the psychometric performance of a Persian translation of the CUDIT-R in a sample of university students from Tehran, Iran, who reported a history of regular cannabis use. We used the well-known standards for medical questionnaires from the Scientific Advisory Committee (SAC) [12]. Also, we looked at evidence that can help prove the CUDIT-R in a psychological setting. This included content-oriented evidence, proof of relationships with conceptually related constructs [13].

2. Methods

2.1. Study Participants

We collected data from university students in Tehran, Iran, in 2024. We recruited respondents through stratified random sampling technique. We contacted respondents through social media platforms such as Telegram, WhatsApp, and

Facebook. This method is most effective when members of the population are difficult to reach, such as people with an addiction and given that cannabis use is illegal in Iran [14].

We first identified a group of university students who were cannabis users, and after collecting data from them, we asked for help identifying other cannabis users. Participants were young university students in Tehran, Iran (ages 19 - 24). In total, we contacted 1350 students to encourage their participation in this cannabis consumption-related study. After communicating about the study and building trust among the participants, 541 students consented to participate in it. We informed them that their personal information would remain confidential due to the illegality of cannabis use in Iran.

Ethical considerations

Before participating, all participants signed the consent letter. This study uses the research ethics committee certificate code of the Najafabad branch of Islamic Azad University, Iran (IR.IAU.NAJAFABAD.REC.1403.097). Iran's cannabis ban guarantees the confidentiality of participants' personal information.

Inclusion criteria:

According to the fifth edition of the Diagnostic and Statistical Mental Health Analysis, Text Revision Criteria (DSM-5-TR), participants must have at least twelve months of experience with marijuana use. They must also avoid stimulant drugs other than cigarettes and marijuana. The participants must be between 18 and 25 years of age and have signed an ethical form indicating their consent.

Exclusion criteria:

Participants who require more advanced care than what an outpatient group can provide, such as those experiencing active mania or psychosis. Any psychiatric or physical condition may find it challenging to communicate effectively. The participant is concurrently receiving another medical or psychological intervention.

2.2. Measures

2.2.1. Cannabis Use Disorders Identification Test-Revised (CUDIT-R)

Researchers have extensively studied the CUDIT-R and found it to be a reliable and valid screening tool for identifying problematic cannabis use worldwide. It has demonstrated good internal consistency, concurrent validity, and discriminant validity among college students [4] [11]. The CUDIT-R consists of eight self-reported items assessing the past six months of cannabis consumption and consequences. We recorded the responses to the items on a five-point Likert scale from 0 to 4, then summed them to obtain a total score [5].

2.2.2. The Structured Clinical Interview for DSM-5 - Clinical Trials Version (SCID-5-CT)

First *et al.* (2016) [15] developed the SCID-5-CT, a structured clinical interview for DSM-5 disorders. It has been widely used in research and clinical settings for evaluation and diagnosis [16]; making it a valuable tool for mental health professionals. Amirinia (2024) [1] used the Persian version of SCID-5-CT on 165

university students who used cannabis, demonstrating high reliability and validity in Iran. The mean and standard deviation (13.83 ± 5.03) were found to have a high omega coefficient and good Richardson Kadre ($\omega = 0.85$; $KR20 = 0.86$). The results also showed enough stability in test-retest with a Kappa index of 0.66 (ICC = 0.66, 95 CI% 0.56, 0.76). We conducted 90-minute interviews with all participants who completed the CUDIT-R-Pr, using the Persian version of SCID-5-CT. As specified in the DSM-5, CUD diagnosis is defined as meeting at least two diagnostic criteria out of 11. “Mild” CUD corresponds to subjects meeting levels two or three. In comparison, “moderate” CUD refers to subjects meeting levels four or five, while “severe” CUD refers to subjects meeting levels six or more.

2.3. Procedure

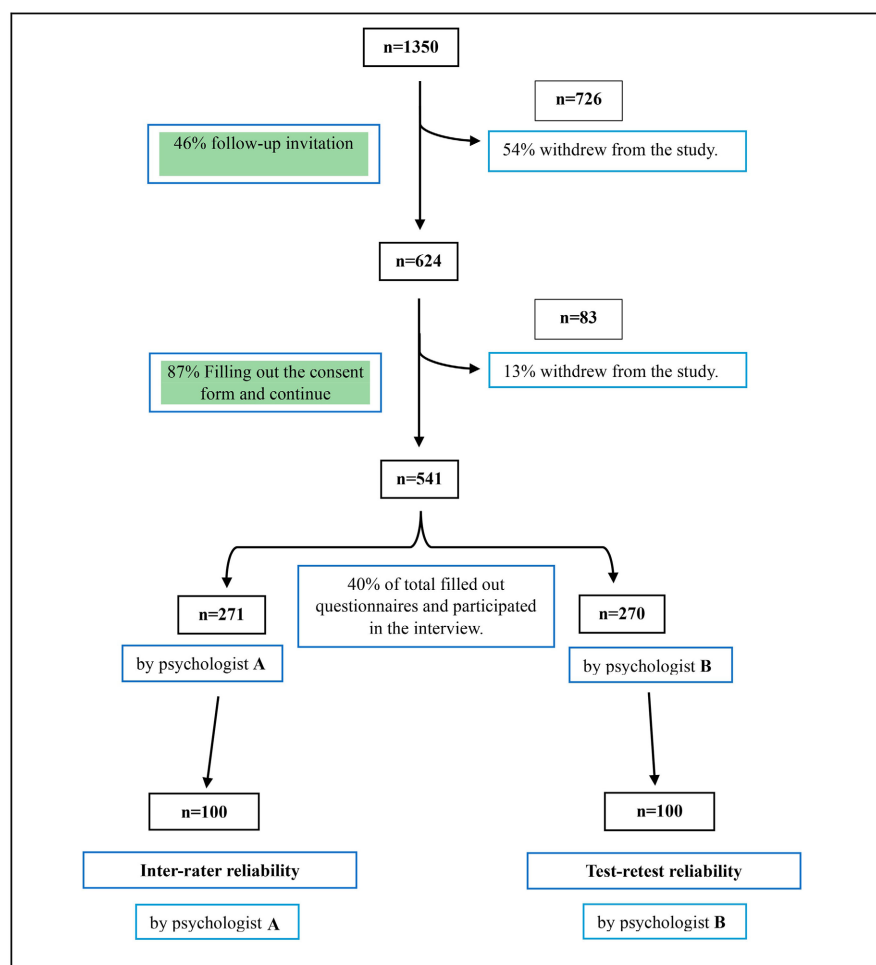


Figure 1. Flowchart of participation.

Two separate psychologists formed two groups, Group A ($n = 271$) and Group B ($n = 270$), and simultaneously interviewed participants in a free 90-minute session based on the DSM-5 criteria for CUD and the Persian version of SCID-5-CT. In the second step, we organized and coordinated a team of two psychologists (A and B) for the equivalence (parallel) method, as Nunnally (1994) [17] supposed, in two

different clinical and social fields to standardize the CUDIT-R. Because a double review by independent reviewers is required to focus on many stages, such as literature search, review process, and quality assessment, to resolve disputes that may arise during the study inclusion process, we also included a third independent reviewer on our team (Figure 1).

2.4. Data Analysis

Boateng *et al.* (2018) [18] recommended a general framework to evaluate the psychometric properties of our proposed scale. We assessed the proposed CUDIT-R Persian version using the criteria below for validity and reliability. We conducted item response theory and ROC analysis in STATA 17 [19], confirmatory factor analysis (CFA) in AMOS 29 [20], and the remaining analyses in IBM SPSS 29 [21].

2.4.1. Translation and Face Validity

Drost (2011) [22] defines face validity as a subjective judgment about the operationalization of a construct. In order to achieve the desired face validity, we revised the item wordings based on expert opinions. Initially, we translated CUDIT-R to Persian. Then, an English-speaking native speaker performed a back translation. In both stages, we received advice from five experts regarding the face validity of the items. After revising the item wordings to solve ambiguities and discussing the visual and qualitative content of the questionnaire, the authors approved the final Persian version of the instrument (Appendix).

1) Content Validity

Bollen (1989) [23] proposed numerous criteria that were used in this investigation. We ensured that 1) the questionnaire aimed to investigate recognized cannabis use; 2) the study's target demographic comprised university students in Iran aged 19 to 24 who used cannabis; and 3) the final wording of the questionnaire items is available in Table 1.

Table 1. Observed distribution of scores for each item of the CUDIT-R, n = 541.

Item	0		1		2		3		4	
	n	%n	n	%n	n	%n	n	%n	n	%n
Frequency of use ^a	0	0.0%	96	65.3%	51	21.9%	1	0.9%	1	2.0%
Hours stoned ^b	0	0.0%	122	50.0%	25	14.6%	2	1.9%	0	0.0%
Unable to stop ^c	0	0.0%	119	64.7%	28	13.5%	2	1.7%	0	0.0%
Fail to do what is expected ^c	1	25.0%	123	41.6%	22	13.8%	3	4.9%	0	0.0%
Time devoted ^c	1	33.3%	120	44.3%	24	13.6%	3	4.2%	1	5.6%
Memory or concentration problems ^c	0	0.0%	112	43.1%	37	18.4%	0	0.0%	0	0.0%
Physically hazardous ^c	0	0.0%	116	41.1%	30	16.3%	3	5.1%	0	0.0%
Cutting down ^d	133	47.2%	-	-	15	8.9%	-	-	1	1.1%

NOTE. Values represent the number of participants with percent (n = 541). ^aResponse options: 0 = never, 1 = monthly or less, 2 = 2 - 4 times a month, 3 = 2 - 3 times a week, 4 = 4 or more times a week. ^bResponse options: 0 ≤ 1, 1 = 1 or 2, 2 = 3 or 4, 3 = 5 or 6, 4 = 7 or more. ^cResponse options: 0 = never, 1 = less than monthly, 2 = monthly, 3 = weekly, 4 = daily or almost daily. ^dResponse options: 0 = never, 2 = yes, but not in the past year, 4 = yes, during the past year.

2.4.2. Construct Validity

The construct validity refers to how our translated CUDIT-R can detect cannabis users in a functioning and operational reality [22]. We conducted confirmatory factor analyses (CFA) to assess the internal consistency and dimensions of the CUDIT-R-Pr. For the CFA calibration sample, we used the principal components method with variable rotation ($n = 541$) to easily pull out one factor from eight CUDIT-R-Pr questions. We employed the Kaiser-Meyer-Olkin (KMO) method to verify the appropriateness of our sample technique for the analysis [24] [25]. According to Sofroniou & Hutcheson (1999) [26] indicated that KMO values exceeding 0.8 were deemed favorable for each item.

2.4.3. Reliability

Ebadi *et al.* (2017) [24] stated reliability as the consistency of scores produced by an instrument, indicating that it yields the same measurements at a specific time or across a duration. Sofroniou & Hutcheson (1999) [26] identified two types of errors in an instrument: 1) systematic or biased errors, which arise in a consistent manner and are assessed through internal consistency, test-retest reliability, and inter-rater reliability; and 2) random errors, which are attributable to unpredictable factors.

1) Internal consistency

Terwee *et al.* (2007) [25] describe internal consistency as a method for assessing reliability, with Cronbach's alpha serving as a measure of internal consistency that evaluates the extent of inter-correlation among scale items.

a) Cronbach's alpha

Cortina (1993) [27] indicated that Cronbach's alpha or McDonald's omega can be used to assess the reliability of item-specific variance in a one-dimensional experiment, particularly when factor-analytic techniques confirm the experiment's one-dimensionality. The CUDIT-R-Pr is an effective tool for measuring fundamental aspects such as low question count, multiple response choices, and unidimensionality. We attempted to conduct the homogeneity indicators as accurately as possible to avoid using Cronbach's alpha as a consistency coefficient, which is not a valid measure

b) Split-half

The split-half approach, another method of test reliability, assumes several items to measure behavior and adjusts the correlation between the two halves of the test to obtain a reliability coefficient for the entire test [22] [23]. Therefore, we conduct these two half-tests concurrently. This allows us to control the effect of memory, which interferes with and limits the test-retest process [22] [23].

2.4.4. Stability over Time (Reproducibility)

Inter-rater reliability, or reproducibility, refers to a test's time consistency from one measurement session to another, also known as test-retest reliability [22]. In this study, we randomly selected one hundred participants from group B after a month, and psychologist B reassessed them. In this way, the correlation between

scores on identical tests at different times operationally defines the reliability of testing and retesting it. ICC is the most suitable and commonly used parameter to assess the reliability of continuous measures after a test-retest by one hundred participants [25]. This period is long enough to prevent recall, and the CUDIT-R is based on the lowest meaningful scale [28]. In other words, we set option 1 to “monthly or less”, making it an excellent point to consider while also ensuring that any changes are unlikely to have occurred. Indeed, we eliminated the individual exhibiting signs of cannabis withdrawal.

1) Inter-rater (observer) reliability

When different raters repeat a measurement under the same conditions and subjects, the degree of agreement is known as inter-rater reliability [29]. In this study, from the beginning, we divided the stages of data collection, implementation, and evaluation of participants between two psychologists: appraiser A for two hundred and seventy-one participants and appraiser B for two hundred and seventy participants. They kept the process a secret from each other until the results analysis was complete. As a result, the psychologist in Group B selected and reassessed one hundred of the two hundred and seventy-one participants. Finally, we used the intra-cluster correlation coefficient (ICC) to measure inter-rater reliability, which is the most suitable and commonly used parameter to assess the reliability of continuous measures [25].

2) Agreement

The agreement and reliability parameters are different and focus on two different questions; the agreement parameter is “How good is the agreement between repeated measurements?”, which expresses measurement error and evaluates precisely how close the scores of repeated measurements are. However, the reliability question asks, “How reliable is the measurement?” [30]. The standard error of measurement (SEM) adequately expresses the measurement error [25]. The SEM equals the square root of the error variance of an ANOVA analysis, either including systematic differences ($SEM_{\text{agreement}}$) or excluding them ($SEM_{\text{consistency}}$) [30].

Terwee (2007) [25] calculated the standard error of measurement (SEM) by taking the square root of the error variance of the ANOVA analysis, either including systematic differences ($SEM_{\text{agreement}}$) or excluding them ($SEM_{\text{consistency}}$). In this study, the $SEM_{\text{agreement}}$ was measured using the $SD \sqrt{1-ICC}$ formula, which was measured for one hundred participants who were individually interviewed by two psychologists ($n = 100$). The SEM was converted into the slightest. The SEM was converted into the slightest detectable change (SDC) using the formula $(1.96 \sqrt{2} SEM)$, (1.96) because of the 95% confidence, and (2) because of the difference of two variances, which reflects the slightest within-person change in score that, with $p < 0.05$, can be interpreted as a real change in one individual ($SDC_{\text{individual}}$) [31]. We used the minimal important change (MIC) value and an anchor-based approach to analyze the smallest score difference in the area of interest that participants found useful.

2.4.5. Responsiveness (Longitudinal Validity)

Researchers define responsiveness as a questionnaire's ability to detect clinically significant changes over time, even if they are minor, and test it by relating the SDC to the MIC [25], recommended an anchor-based method to determine the MIC because distribution-based methods do not provide a good indication of the importance of the observed change [31]. Another adequate measure of responsiveness is the area under the receiver operating characteristics (ROC) curve (AUC) which is a measure of the questionnaire's ability to differentiate participants who have or have not changed according to an external criterion [30].

2.4.6. Interpretability (Internal Validity)

When we can assign qualitative meaning to quantitative scores, it means that the questionnaire has interpretability [30]. For interpretability, we provide the means and standard deviations of the participants' scores before and after the test [12]. We analyzed the MIC using an anchor-based approach and decided to use the ability item (cannabis use rate) as an anchor question [31]. Students who performed significantly better on the ability item in the test showed a corresponding change. Based on changes in the CUDIT-R-Pr scores, the Pearson correlation coefficient confirmed the usefulness of the anchor question. The correlation coefficient (-0.21) should be considered a favorable anchor.

2.4.7. Floor and Ceiling Effects

When measuring cannabis use, one important thing to think about is how sensitive the CUDIT-R-Pr is to changes in cannabis use at the lower and upper ends of the scale. If floor or ceiling effects are present, it's likely that extreme items are missing at the lower or upper end of the scale, which means the scale doesn't have a lot of content validity [32]. We assessed the presence of ceiling and floor effects based on the percentage of patients with the highest or lowest CUDIT-R-Pr scores, considering whether this was the case for 15% or more of the patients [25].

2.4.8. Criterion Validity

Criteria validity refers to the degree of correspondence between a test measure and one or more external criteria, usually measured by their correlation [22]. Some researchers have described a gold standard based on criterion validity [25], but related, constructed, and currently available standard CUD questionnaires are unavailable. We also opted to solely utilize the CUDIT-R-Pr, incorporating the DSM-5 criteria for CUD and calculated its sensitivity and specificity. Mean CUDIT-R-Pr scores significantly differed across the four DSM-5 diagnostic severity levels (Figure 2).

1) Diagnostic test analysis

In order to compare the performance of CUDIT-R with other diagnostic tests, we applied the ROC curve to interpret the undersurface area (AUC) and determined the optimal cutting point, which is a critical component of these analyses. Over the past decade, significant research on CUDIT-R has focused on determining the sensitivity and specificity of this questionnaire, as well as determining its

cutoff point, using AUC as a measure of discriminatory ability [4]-[7] [9]-[11] [33]. We determined the sensitivity and specificity of the CUDIT-R to distinguish between individuals who met the criteria for CUD as positive cases and those who did not as negative cases, based on the measurements made using the SCID-5-CT [11] [28] by STATA 17.0 [19]. We achieved a cutoff of ten and less for CUDIT-R-Pr to predict any DSM-5, based on the exact maximum correctly classified value (0.887) with high levels of sensitivity (0.96), specificity (0.69), and Youden value (0.65). Similarly, we achieved a cutoff of twelve and less for CUDIT-R-Pr to predict moderate DSM-5, based on the exact maximum Youden index (0.72) with high levels of sensitivity (0.82), specificity (0.89), and correctly classified (0.84) [11].

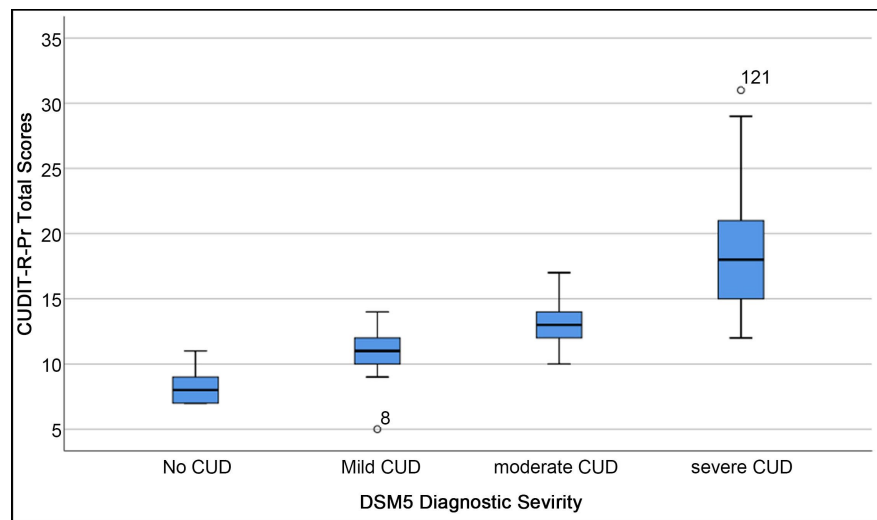


Figure 2. CUDIT-R-Pr distributions across DSM 5 severity categories.

2) Discriminant validity

We also examined the psychometric properties of individual CUDIT-R items through an item response theory (IRT) analysis [34] [35]. IRT is a statistical method that models the interaction between a subject's ability and the difficulty of test items. It is particularly useful in developing and scoring multi-item scales [34]. IRT estimates a parameter for each item and participant, and can be used to measure underlying latent traits [35]. In addition to implementing IRT for CUDIT-R-Pr, we compared the results and outputs of IRT with similar research in CUDIT-R [4] [11]. Descriptions of items in the GR model are described by an alpha discrimination parameter (α) and a category threshold or location parameter (β) [11].

3. Results

3.1. Participants' Characteristics and Response Rates

The median (M) and mean standard deviation (SD) describe age, sex, the number of DSM-5 criteria, and the CUDIT-R-Pr item distributions, while the number and

percentage (n%) are as follows.

We selected 541 single and undergraduate participants (ages 19 to 24), with a gender distribution of men (53.6%) and females (46.4%). One hundred seventy-two participants (21.6%) did not meet the DSM-5 criteria for CUD; 82 participants (15.2%) met two or three criteria for CUD (mild CUD); 124 participants (22.9%) met four or five criteria for CUD (moderate CUD); and 280 participants (40.3%) met six or more criteria for CUD, which was consistent with severe CUD (Table 2).

Table 2. Distribution, means, and standard deviation of sample characteristics, n = 541.

		n	%n	M	SD	Variance	Median	Min	Max
Gender	Male	290	53.6	13.96	5.18	26.82	13	5	31
	Female	251	46.4	13.69	4.87	23.69	13	7	29
Age	19 - 20	58	10.7	13.81	5.54	30.65	12	7	31
	20 - 21	91	16.8	14.02	5.08	25.78	13	5	26
	21 - 22	164	30.3	13.85	4.95	24.48	13	7	27
	22 - 23	142	26.2	13.47	4.67	21.77	13	7	28
	23 - 24	86	15.9	14.20	5.45	29.70	13	7	29
DSM-5 scores	No diagnosis	117	21.6	8.15	1.07	1.15	8	7	11
	Mild	82	15.2	10.74	1.39	1.95	11	5	14
	Moderate	124	22.9	12.87	1.67	2.80	13	10	17
	Severe	218	40.3	18.58	3.99	15.94	18	12	31

Regarding the Persian translation of the CUDIT-R (CUDIT-R-Pr), the one-factor solution presented a satisfactory fit, in line with the initial validation of the scale in English. CUDIT-R-Pr scores increased with the severity of the DSM-5 diagnostic levels, with mean scores ranging from 8.80 to 18.86. The mean CUDIT-R-Pr scores significantly differed across the four DSM-5 diagnostic severity levels.

The mean and standard deviation numbers of the DSM-5 criteria were 2.82 and 1.18, respectively. The mean and standard deviation of the CUDIT-R-Pr total score were, respectively, 13.82 and 5.03 (SD = 9.05). The CUD rate varied depending on the instrument used. 70.4 percent met the DSM-5 criteria for CUD (n = 150); 78.4 percent of participants (n = 424) were screened positive based on the CUDIT-R-Pr with adequate ICC in our sample (KR20 = 0.86; ω = 0.85) (Table 3).

Table 3. Descriptive statistics and reliability coefficient among study variables, n = 541.

	M	SD	Skewness	Kurtosis	Reliability coefficient
DSM-5 based on SCID-5-CT	2.82	1.18	-0.44	-1.33	KR20 = 0.86; ω = 0.85
CUDIT-R-Pr	13.83	5.03	0.78	0.14	α = 0.76; ω = 0.82

NOTE. 0 = no diagnosis (0 - 1 symptoms), 2 = mild (2 - 3 symptoms), 3 = moderate (4 - 5 symptoms), and 4 = severe (6 or more symptoms).

3.2. Factorial Structure

We performed CFA to test the proposed CUDIT-Pr scale's uni-dimensionality. We specified a latent variable model in which all eight indicators (items) were loaded onto a single latent factor. The model showed an acceptable fit to the data (χ^2 (df) = 84.829 (20), CFI = 0.947, TLI = 0.926, RMSEA [90% CI] = 0.077 [0.061, 0.095]) and all standardized factor loadings ranged between 0.52 - 0.73 (Figure 3).

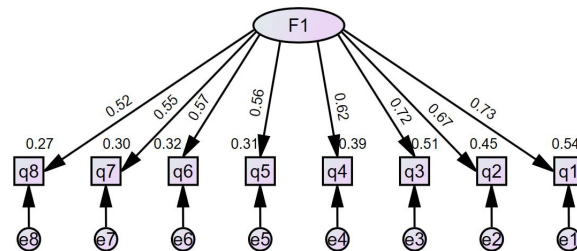


Figure 3. Confirmatory factor analysis results.

3.3. Internal Consistency

The Cronbach's alpha was calculated for each explicit question, ranging from 0.72 to 0.81, and for the total CUDIT-R-Pr, it was 0.76 (ranging from 0.72 to 0.81), as shown in Table 4. Additionally, McDonald's Omega was 0.82 [36].

Table 4. The Corrected Item Total Correlation (CITC) and cronbach's alpha of the CUDIT-R-Pr.

Item	All participants (n = 541)				Test-retest (n = 100)			
	M	Variance	CITC	α	M	Variance	CITC	α
Frequency of use ^a	25.54	88.523	0.700	0.732	25.30	65.525	0.582	0.701
Hours stoned ^b	25.87	90.032	0.633	0.738	25.71	65.966	0.608	0.702
Unable to stop ^c	25.68	88.982	0.683	0.734	25.66	66.469	0.514	0.708
Fail to do what is expected ^c	26.03	91.038	0.598	0.742	26.11	68.968	0.425	0.719
Time devoted ^c	25.97	91.597	0.569	0.744	26.00	67.495	0.527	0.711
Memory or concentration problems ^c	25.97	91.969	0.579	0.745	25.89	68.907	0.387	0.721
Physically hazardous ^c	26.02	91.920	0.584	0.745	26.05	67.321	0.528	0.710
Cutting down ^d	26.37	82.693	0.601	0.723	26.28	63.355	0.421	0.709
Total	13.83	25.338	1.000	0.814	13.80	18.808	1.000	0.694

Split-half

The Cronbach's alpha, calculated using the split-half method, was 0.80 for part 1 and 0.59 for part 2. The Spearman-Brown coefficient, both at the same equal and unequal length, was 0.91, while the Gutman coefficient was 0.76 [26].

Inter-observer reliability

The A two-way random-effects model calculated the ICC for the total scores on

the questionnaire, which ranged from a minimum score of 0 to a maximum score of 32. Researchers recommend an ICC higher than 0.70 as a minimum standard for reliability [30]. At this point, a psychologist from group B simultaneously reassessed 100 participants from group A using the DSM5 for CUD and the CUDIT-R-Pr test ($n = 100$, M difference = -0.81 , SD difference = 2.48). The CUDIT-R-Pr's ICC was high (ICC = $0.86.95$, CI $0.79 - 0.91$), indicating a high degree of reliability and higher than the threshold of 0.7 . If Kendall's tau-b, which measures the relationship between the difference in the sum of the scores and the difference in their means, was not significant (p value = $0.08 > 0.05$), we knew that there was no systematic error [29]. The measure of agreement with the Kappa index was 78% , which was substantial agreement [37].

Inter-rater (observer) reliability

Additionally, two psychologists conducted a parallel measurement of Cronbach's alpha between groups A ($n = 271$) and B ($n = 270$), yielding an adequate coefficient of 0.84 . We calculated the ICC for the total scores on the questionnaire, which ranged from a minimum score of 0 to a maximum score of 32, using a one-way random-effects model. The ICC was high (0.84 , 95% CI = $0.80, 0.88$), indicating a high degree of reliability and higher than the threshold of 0.7 [30].

Reproducibility

The same psychologist randomly selected 100 participants from group B for this study, and after a month, they reassessed and completed the CUDIT-R-Pr. We calculated the ICC for the total scores on the questionnaire, which ranged from a minimum score of 0 to a maximum score of 32, using a two-way random-effects model [25]: ($n = 100$, M difference = -0.17 , SD difference = 1.70). The CUDIT-R-Pr's ICC was high (ICC = 0.91 , 95% CI $0.86 - 0.94$), indicating high reliability and higher than the threshold of 0.7 . We found no systematic error because Kendall's tau-b, which measures the relationship between the difference in the sum of the scores and the difference in their means, was not significant (p value = $0.51 > 0.05$). The agreement estimate is subjective and does not provide an objective index such as the ICC. The agreement estimate is subjective and does not provide an objective index such as the ICC [24] [37]. So, we have to do more evaluation.

Agreement

We calculated the $SDC_{\text{individual}} = 2.57$ ($n = 100$) and found that the $SEM_{\text{agreement}}$ was 0.93 . Additionally, we can measure SDC in a group of individuals, known as the SDC_{group} , by dividing the $SDC_{\text{individual}}$ by \sqrt{n} ($SDC_{\text{group}} = 0.26$; $n = 100$) [12] [25].

Responsiveness

The $SDC_{\text{individual}}$ and SDC_{group} were 2.57 and 0.25 smaller than the MIC (3), indicating a responsive CUDIT-R-Pr [31].

Interpretability

Students who scored "much better" or "better" on the satisfaction statement saw improvements from a mean score of 12 to 15 and 13.40 to 16.40 , respectively (3). For students who were neither "better" nor "worse", there was some

improvement; this was from 11.13 to 11.35 (0.22). For students who were “worse” or “much worse”, the mean scores decreased from 14.40 to 11 (a decrease of 3.40) and from 14.50 to 11.50 (a decrease of 3) (Table 5).

Table 5. Distribution of based on an anchor question, n = 100.

How much have you changed in the last month?	Mean (SD) just before test	Mean (SD) after a month test	Change in score
Much better	12.00 (2.00)	15.00 (2.00)	+3.00
Better	13.40 (1.44)	16.40 (1.44)	+3.00
No change	11.13 (0.31)	11.35 (0.29)	+0.22
Worse	14.40 (1.21)	11.00 (1.18)	-3.40
Much worse	14.50 (1.50)	11.50 (2.50)	-3.00
MIC	-	-	3

Floor and Ceiling Effects

We observed floor effects in less than 7 percent of cases at the CUDIT-R-Pr test (valid n = 100) and less than 2 percent at the retest (valid n = 100). We observed ceiling effects in less than 8 percent of cases at the CUDIT-R-Pr test and 2 percent at the retest. Ceiling or floor effects were considered not present, as the percentages did not exceed 15 percent.

3.4. Measurement Invariance

We conducted measurement invariance testing on two groups of respondents (*i.e.*, females, n = 251 and males, n = 290) to assess whether the measurement model's psychometric properties can be generalized by sex [38]. As presented in Table 6, results indicated configural, factorial, and scalar invariance of our measurement model between the female and male sub-samples. First, we estimated the unconstrained measurement models simultaneously across the two sub-samples. The goodness of fit indices displayed a good fit of the models to the data and supported configural invariance (Table 6). The fit statistics of the unconstrained models served as the baseline for comparing the fit of the following constrained models, where, in each step, we increasingly imposed model constraints where parameters were set to be equal across female and male sub-samples.

Table 6. Measurement invariance testing results between female (n = 251) and male (n = 290) sub-samples.

Model	χ^2 (df)	CFI	TLI	RMSEA [90% CI]	$\Delta\chi^2$ (Δ df)	p-value
Unconstrained (configural invariance)	102.825 (40)	0.949	0.928	0.054 [0.041; 0.067]	-	-
Measurement weights constrained (factorial invariance)	110.582 (47)	0.948	0.938	0.050 [0.038; 0.062]	7.757 (7)	0.645
Measurement intercepts constrained (scalar invariance)	117.890 (55)	0.949	0.948	0.046 [0.035; 0.058]	15.065 (15)	0.553

Next, we imposed constraints on measurement weights between the two sub-samples to assess factorial invariance, which resulted in no significant decline in model fit compared to the unconstrained model ($\Delta\chi^2$ (Δdf) = 7.757 (7), p -value = 0.645). In the last step, we constrained the measurement intercepts to be equal between female and male sub-samples for testing scalar invariance. This constraint did not cause a significant decline in model fit compared to the baseline model ($\Delta\chi^2$ (Δdf) = 15.065 (15), p -value = 0.553). Thus, we concluded invariance of the CUDIT-Pr models between the female and male sub-samples suggesting that the performance of the CUDIT-Pr scale does not vary by sex.

3.5. Discriminant Validity

ROC Analyses

All pairwise comparisons were significant ($p < 0.001$), with a Pearson correlation of 0.81. Looking at the ROC curve with a self-reported CUD as the criterion showed a decent fit (AUC = 0.95, CI: 0.92 - 0.96, $p < 0.001$) and a standard error of 0.01 (Figure 4).

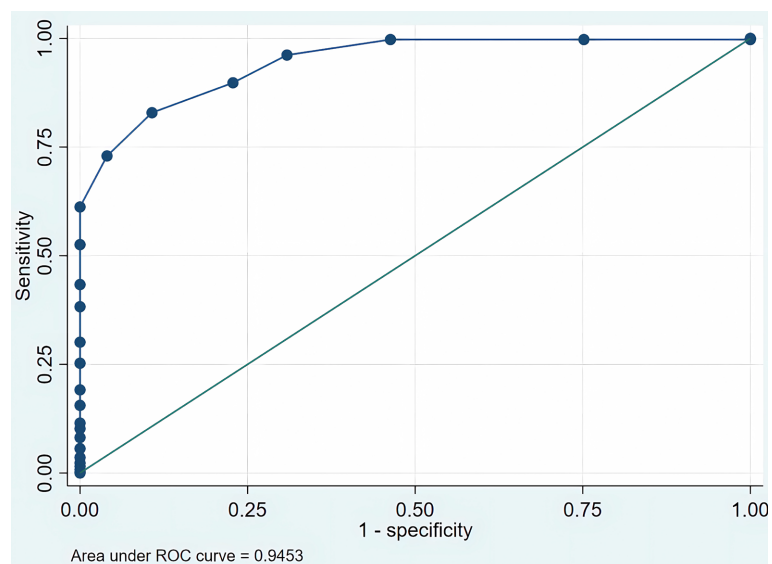


Figure 4. ROC curve showing the optimal cutoff of 12 for CUDIT-R-Pr.

Sensitivity and Specificity

We found that a cut-point of ten or less was best for predicting any DSM-5, with sensitivity, specificity, and the Youden index at 96.17 percent, 69.13 percent, and 0.65 percent, respectively. This was based on the exact highest correctly classified value (88.72%), with LR+ (3.11) for a positive result and LR- (0.05) for a negative result. In addition, we calculated the optimal cut-point of twelve and less, taking into account the exact maximum Youden index (0.72), the likelihood ratio for a positive result (LR+), and the likelihood ratio for a negative result (LR-), which are respectively 7.72 and 0.19. The ROC curve at each CUDIT-R-Pr cut point is shown in Table 7.

Table 7. Identification of current self-reported CUD by Persian version of CUDIT-R (CUDIT-R-Pr), n = 541.

Cutoff point	Sensitivity	Specificity	Correctly classified	Youden Index	LR+	LR-
≥5	100.00%	0.00%	72.46%	0.00	1.0000	0.0103
≥7	99.74%	0.00%	72.27%	0.00	0.9974	0.0103
≥8	99.74%	24.83%	79.11%	0.25	1.3270	0.0103
≥9	99.74%	53.69%	87.06%	0.53	2.1539	0.0048
≥10 ^a	96.17%	69.13%	88.72%	0.65	3.1152	0.0554
≥11	89.80%	77.18%	86.32%	0.67	3.9352	0.1322
≥12 ^b	82.91%	89.26%	84.66%	0.72	7.7208	0.1915
≥13	72.96%	95.97%	79.30%	0.69	18.1182	0.2818
≥14	61.22%	100.00%	71.90%	0.61	0.0103	0.3878
≥15	52.55%	100.00%	65.62%	0.53	0.0103	0.4745
≥16	43.37%	100.00%	58.96%	0.43	0.0103	0.5663
≥17	38.27%	100.00%	55.27%	0.38	0.0103	0.6173
≥18	30.10%	100.00%	49.35%	0.30	0.0103	0.6990
≥19	25.26%	100.00%	45.84%	0.25	0.0103	0.7474
≥20	19.13%	100.00%	41.40%	0.19	0.0103	0.8087
≥21	15.56%	100.00%	38.82%	0.16	0.0103	0.8444
≥22	11.48%	100.00%	35.86%	0.11	0.0103	0.8852
≥23	10.20%	100.00%	34.94%	0.10	0.0103	0.8980
≥24	8.16%	100.00%	33.46%	0.08	0.0103	0.9184
≥25	5.61%	100.00%	31.61%	0.06	0.0103	0.9439
≥26	3.57%	100.00%	30.13%	0.04	0.0103	0.9643
≥27	2.30%	100.00%	29.21%	0.02	0.0103	0.9770
≥28	1.53%	100.00%	28.65%	0.02	0.0103	0.9847
≥29	0.77%	100.00%	28.10%	0.01	0.0103	0.9923
≥30	0.26%	100.00%	27.73%	0.00	0.0103	0.9974
≥31	0.00%	100.00%	27.54%	0.00	0.0103	1.0000

^aOptimal cutoff point for CUDIT-R-Pr to predict any DSM-5 based on the exact maximum correctly classified value. ^bOptimal cutoff point for CUDIT-R-Pr to predict moderate DSM-5 based on exact maximum Youden Index.

3.6. Item Characteristics

The IRT result demonstrated an S-shaped curve that described the relationship between the probability of a correct response to an item and the ability scale, known as the item characteristic curve [31]. In this study, the probability was slightly higher than the 10 score (Figure 5).

The item parameters α and β estimated from the GR model demonstrated that all questions had a high discrimination in the range of 1.26 to 2.20 in the alpha parameter. Of course, Item 1 was the highest ($\alpha = 2.20$), and Item 8 was the lowest

($\alpha = 1.26$). Location parameters spanned the broadest range for item 7 ($\beta = -5.35 - 3.29$), item 6 ($\beta = -5.05 - 3.16$), and item 5 ($\beta = -4.39 - 3.03$), in that order. These parameters provided information on both low and high levels of the latent hazardous use construct: Item 4 ($\beta = -3.78 - 2.70$); item 2 ($\beta = -3.55 - 2.61$); and item 3 ($\beta = -3.64 - 2.10$). In order to provide information at both low and high levels of the latent hazardous use construct, item 8 ($\beta = 0.66 - 1.61$) spanned the least range of the latent hazardous use construct. Item 1 ($\beta = -0.77 - 1.75$) exhibited relatively low item difficulty in the sample, although it provided narrower coverage relative to item 8. **Table 8** and **Figure 6** display all estimated item parameters from the GR model and the characteristic curves for Items 1 to 8.

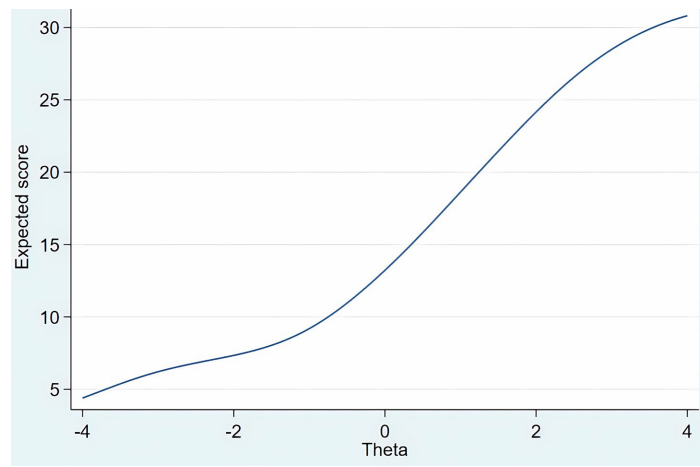
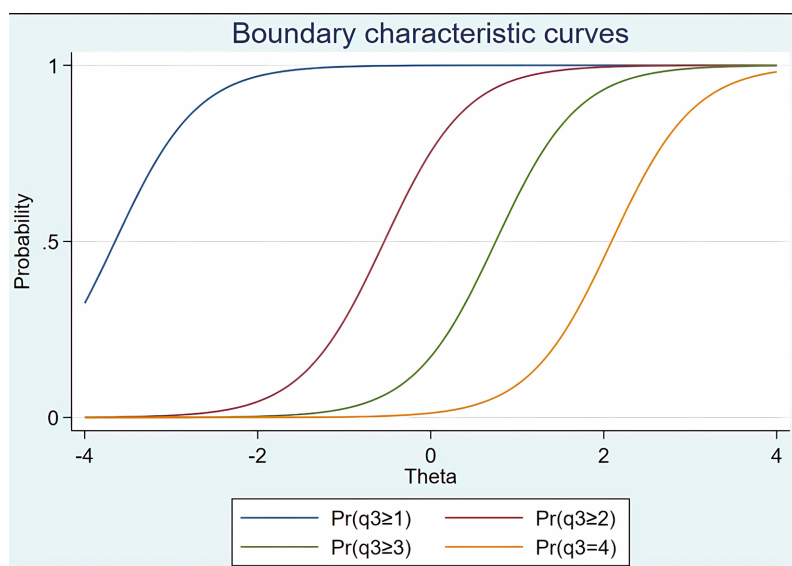
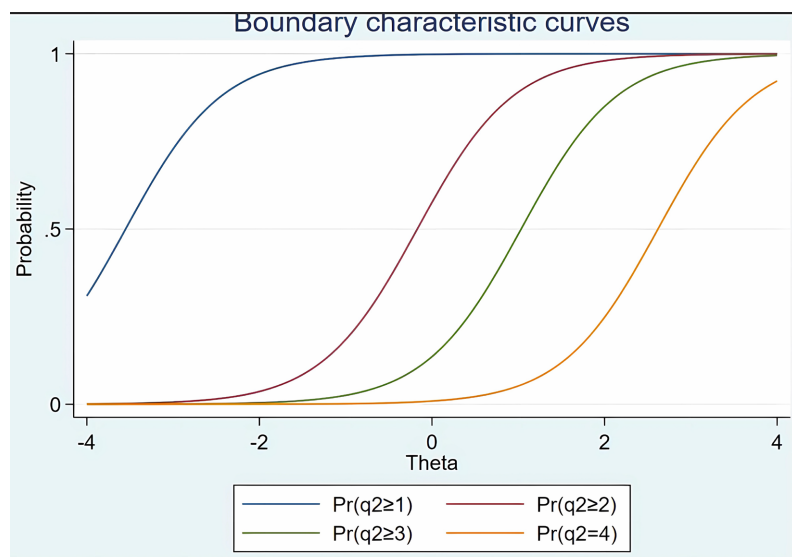
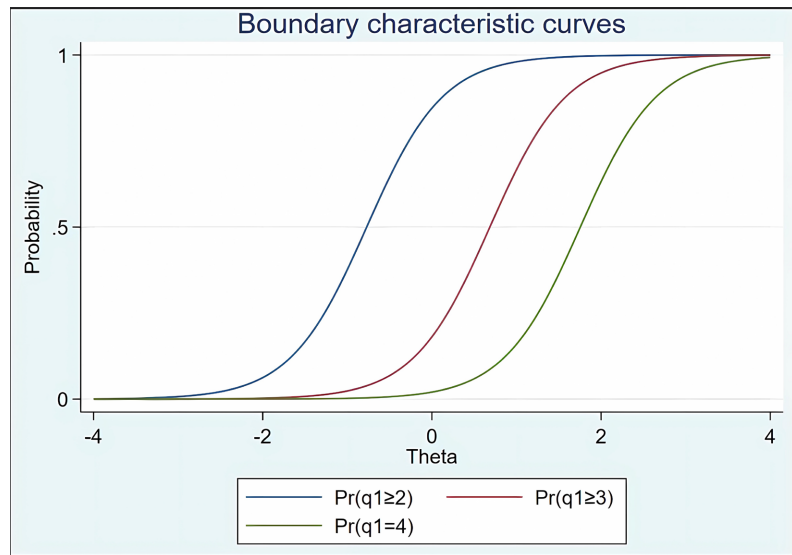


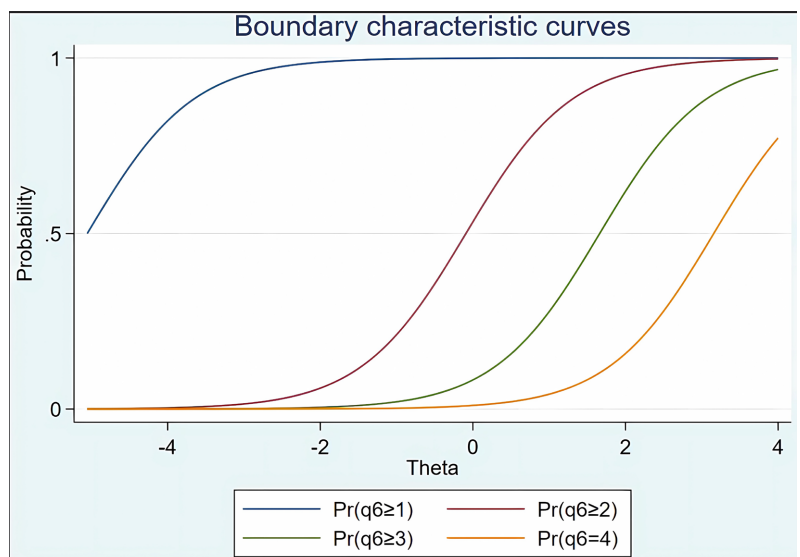
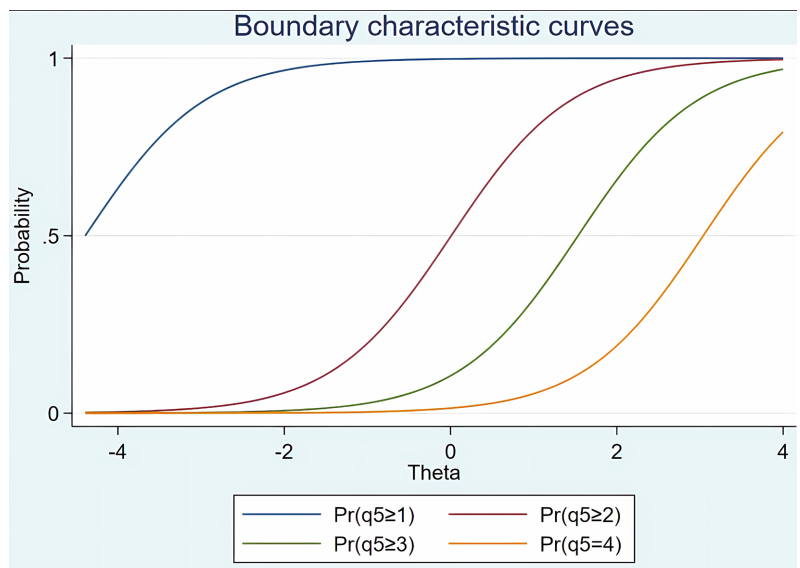
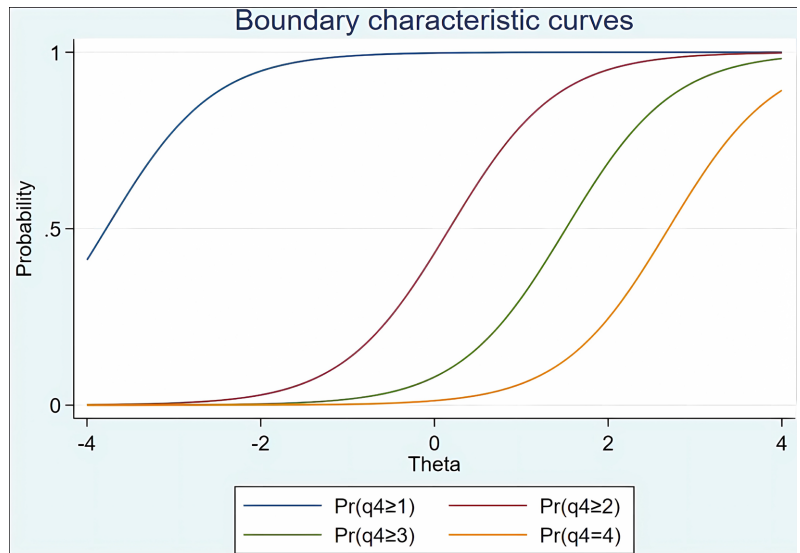
Figure 5. Four Item characteristic curve of CUDT-R-Pr.

Table 8. CUDIT-R item slope (α) and category threshold (β) parameter estimates from the graded response model, n = 541.

Item	Slope α		Threshold β_1		Threshold β_2		Threshold β_3		Threshold β_4	
	Coef.	%95 CI	Coef.	%95 CI	Coef.	%95 CI	Coef.	%95 CI	Coef.	%95 CI
Frequency of use	2.20	1.82, 2.59	-	-	-0.77	-0.93, -0.61	0.69	0.53, 0.84	1.75	1.51, 1.10
Hours stoned	1.79	1.47, 2.11	-3.55	-4.26, -2.83	-0.17	-0.32, -0.03	1.03	0.84, 1.22	2.61	2.21, 3.02
Unable to stop	2.08	1.73, 2.44	-3.64	-4.46, -2.83	-0.54	-0.68, -0.39	0.75	0.59, 0.91	2.10	1.80, 2.38
Fail to do what is expected	1.61	1.32, 1.91	-3.78	-4.57, -2.98	0.17	0.21, 0.32	1.51	1.26, 1.76	2.70	2.25, 3.13
Time devoted	1.39	1.13, 1.66	-4.39	-5.44, -3.35	0.01	-0.15, 0.17	1.53	1.25, 1.81	3.03	2.48, 3.60
Memory or concentration problem	1.44	1.17, 1.72	-5.05	-6.60, -3.50	-0.09	-0.25, 0.07	1.66	1.37, 1.95	3.16	2.60, 3.72
Physically hazardous	1.34	1.07, 1.60	-5.35	-7.06, -3.66	0.07	-0.09, 0.25	1.80	1.48, 2.13	3.29*	2.65, 3.91
Cutting down	1.26	1.00, 1.52	-	-	0.66	-0.11, 0.24	-	-	1.61*	1.29, 1.94

NOTE. α represents the discrimination parameter for each item, or a slope indicating an item’s ability to discriminate between individuals with different levels of the latent hazardous use construct. Each β represents a location parameter, or the point along the latent hazardous use construct at which the probability of a response lying at or above a given category threshold is 0.50. The 95% confidence intervals (CI) represent the higher slope parameter α , the better the item discriminates the latent trait and between-category threshold of the parameters β represent the points along the latent trait at which the probability of responding to a certain category passes 0.50. The greater the variety of threshold parameters β (negative to positive), the more descriptive the item is. Descriptions adapted from [4] [7] [11].





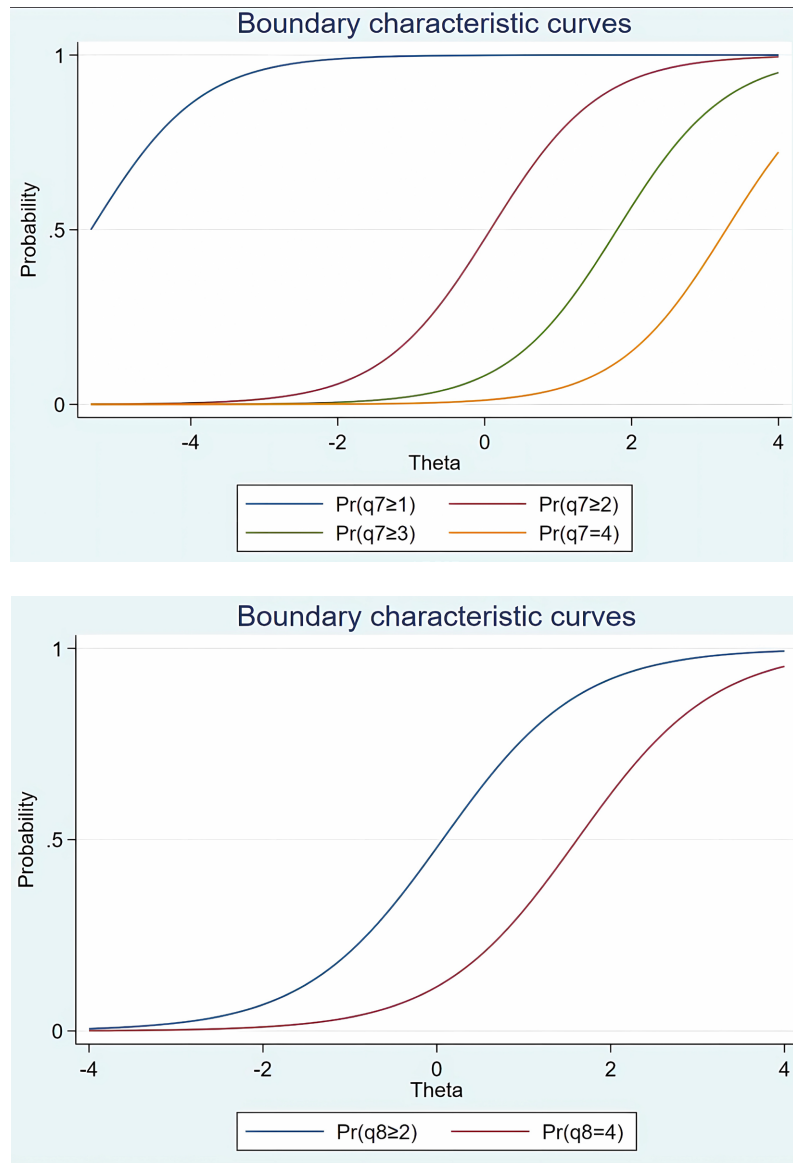


Figure 6. Item (1 to 8) characteristic curves of CUDT-R-Pr by IRT.

4. Discussion

Two groups reviewed the construct validity of the CUDIT-R-Pr: 1) translation validity, which includes face validity and content validity; and 2) criterion-related validity, which relies on discriminant validity. The CUDIT-R-Pr exhibited a high level of accuracy in distinguishing between young adult students with cannabis use. A correctly classified value of 88.72% indicates a cutoff of ten or less at universities with and without CUD. Given that the Youden index is high (0.72) at cutting points 12 and below, we could conclude that this is the optimal location for identifying the moderate level of CUD based on the DSM-5. Our comparative reviews support our guess that we will find a severe level of CUD from the cutoff point 12 up.

In the past, researchers such as Schultz (2019) [4] and Coelho (2024) [11],

found cutoffs of 6 and 9 for CUDIT-R, respectively, in the original English version of CUDIT-R in the adult student population [4] [11]. This study aligns with the findings of the student population, specifically the cut-point of ten or less, which is less than ten. Researchers from the clinical population [5] and the medical and non-medical communities [9] have discovered a threshold of more than ten.

The IRT analysis of the CUDIT-R-Pr performance in a young group of Iranian university students revealed that in items 2 and 4, the number of hours spent high on cannabis and failure to meet expected cannabis use was moderate, indicating a fair amount of trouble for the sample. Furthermore, Coelho (2024) [11] and Schultz (2019) [4] found that item 3's assessment of the inability to stop cannabis use was maximally discriminatory and moderately difficult.

The study's results indicate that Items 7, which evaluate use with potential for physical activity, had the lowest discrimination and the greatest difficulty. According to Coelho (2024) [11], this was likely because the low-frequency confirmation of the high response option made it difficult to stop consumption. Also, Items 6 and 5, which measure time spent and memory or concentration issues, did not do a satisfactory job of telling the difference between the different levels of the latent harmful use construct and high difficulty. This suggests that Items 6 are similar to the findings of Annaheim (2010) [7]. Items 1 and 8, which assessed the frequency of use and attempts to reduce it, were the least difficult, similar to Annaheim (2010) [7] and Schultz (2019) [4] and Coelho (2024) [11]. Although item 1 demonstrated the most discrimination and item 8 the least discrimination, The lack of similarity in discrimination between Item 1 and other research likely stems from the different methods used to select samples from the young student population.

5. Conclusions

In this study, we used typical methods to estimate the test reliability of CUDIT-R-Pr, and after translating and preparing the questionnaire, we assessed its validity. The results of this study, obtained through CFA, confirmed the hypothesis that the CUDIT-R-Pr was unidimensional and demonstrated sufficient reliability. The Cronbach's alpha and McDonald's omega coefficients for the CUDIT-R-Pr were satisfactory enough. Additionally, Cronbach's alpha and McDonald's omega coefficients for the CUDIT-R-Pr were both satisfactory. The inter-rater technique yielded a high ICC value for this test, indicating its high reliability.

The university student's diagnostic severity increased the CUDIT-R-Pr scores, and the ROC curve analysis, which included self-reported CUD criteria (mild, moderate, or severe), showed a great fit, which is in line with other research. The university students' CUDIT-R-Pr scores went up as their diagnosis got worse, and an analysis of the ROC curve using self-reported CUD criteria showed a great fit, which is in line with previous research [4] [11]. As a result, the CUDIT-R-Pr had a higher level of internal consistency, confirming the CUDIT-R-Pr's exceptional stability over time. Therefore, we can conclude that the CUDIT-R-Pr exhibits

reliability, reproducibility, responsiveness, substantial agreement, and interpretability.

Limitation

The main limitation is probably that the assessment of the DSM-5 criteria for CUD was not based on another valid structural questionnaire for cannabis use, which is why we used the factor analysis method for construct validity. However, the main findings of this study on CUDIT-R-Pr were relatively similar to those from the initial validation of CUDIT-R.

Another limitation was that university students who participated in the study should abstain from cannabis use 48 hours before evaluation, potentially excluding those at risk of developing withdrawal symptoms. The exclusion criteria limited the generalizability of the findings, particularly due to the illegality of cannabis use in Iran.

Suggestion

Researchers could use a structured diagnostic tool to compare the CUDIT-R-Pr with another valid questionnaire in Iran. They could look at DSM-5 criteria for CUD, or addictive cannabis disorders, and look at lower screening scores in the student population, the clinical population of cannabis users, and the non-student population.

Credit Authorship Contribution Statement

Mahmood Amirinia: conceptualization, formal analysis, original draft writing, and funding acquisition. **Benjamin Ghasemi:** methodology, formal analysis, writing, review, and editing. **Parisa Aghazadeh:** implementation and evaluation of tests. **Alireza Mollazadeh:** Supervisor.

Acknowledgements

The authors would like to thank all the participants in the study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Amirinia, M. (2024) The Effectiveness of Matrix Treatment on Craving and Psychological Capital in Adults with Cannabis Use Disorder, in Faculty of Medicine Sciences. West Tehran Branch of Islamic Azad University, 133.
- [2] Rastegari, A., Baneshi, M.R., Hajebi, A., Noroozi, A., Karamouzian, M., Shokoohi, M., *et al.* (2022) Population Size Estimation of People Who Use Illicit Drugs and Alcohol in Iran (2015-2016). *International Journal of Health Policy and Management*, **12**, 1-7. <https://doi.org/10.34172/ijhpm.2022.6578>
- [3] Rostam-Abadi, Y., Gholami, J., Jobehdar, M.M., Ardeshir, M., Aghaei, A.M., Olamazadeh,

- S., et al. (2023) Drug Use, Drug Use Disorders, and Treatment Services in the Eastern Mediterranean Region: A Systematic Review. *The Lancet Psychiatry*, **10**, 282-295. [https://doi.org/10.1016/s2215-0366\(22\)00435-7](https://doi.org/10.1016/s2215-0366(22)00435-7)
- [4] Schultz, N.R., Bassett, D.T., Messina, B.G. and Correia, C.J. (2019) Evaluation of the Psychometric Properties of the Cannabis Use Disorders Identification Test—Revised among College Students. *Addictive Behaviors*, **95**, 11-15. <https://doi.org/10.1016/j.addbeh.2019.02.016>
- [5] Adamson, S.J., Kay-Lambkin, F.J., Baker, A.L., Lewin, T.J., Thornton, L., Kelly, B.J., et al. (2010) An Improved Brief Measure of Cannabis Misuse: The Cannabis Use Disorders Identification Test-Revised (CUDIT-R). *Drug and Alcohol Dependence*, **110**, 137-143. <https://doi.org/10.1016/j.drugalcdep.2010.02.017>
- [6] Loflin, M., Babson, K., Browne, K. and Bonn-Miller, M. (2017) Assessment of the Validity of the CUDIT-R in a Subpopulation of Cannabis Users. *The American Journal of Drug and Alcohol Abuse*, **44**, 19-23. <https://doi.org/10.1080/00952990.2017.1376677>
- [7] Annaheim, B., Scotto, T.J. and Gmel, G. (2010) Revising the Cannabis Use Disorders Identification Test (CUDIT) by Means of Item Response Theory. *International Journal of Methods in Psychiatric Research*, **19**, 142-155. <https://doi.org/10.1002/mpr.308>
- [8] Marshall, S.E. (2013) The Cannabis Use Disorder Identification Test-Revised (CUDIT-R): Categorisation and Interpretation. Ph.D. Thesis, University of Tasmania.
- [9] Bonn-Miller, M.O., Heinz, A.J., Smith, E.V., Bruno, R. and Adamson, S. (2016) Preliminary Development of a Brief Cannabis Use Disorder Screening Tool: The Cannabis Use Disorder Identification Test Short-form. *Cannabis and Cannabinoid Research*, **1**, 252-261. <https://doi.org/10.1089/can.2016.0022>
- [10] Risi, M., Sokolovsky, A., White, H. and Jackson, K. (2020) Factor Structure of the Cannabis Use Disorders Identification Test Revised (CUDIT-R) for Men and Women. *Cannabis*, **3**, 148-156. <https://doi.org/10.26828/cannabis.2020.02.002>
- [11] Coelho, S.G., Hendershot, C.S., Quilty, L.C. and Wardell, J.D. (2024) Screening for Cannabis Use Disorder among Young Adults: Sensitivity, Specificity, and Item-Level Performance of the Cannabis Use Disorders Identification Test—Revised. *Addictive Behaviors*, **148**, Article ID: 107859. <https://doi.org/10.1016/j.addbeh.2023.107859>
- [12] Kievit, A.J., Kuijer, P.P.F.M., Kievit, R.A., Sierevelt, I.N., Blankevoort, L. and Frings-Dresen, M.H.W. (2014) A Reliable, Valid and Responsive Questionnaire to Score the Impact of Knee Complaints on Work Following Total Knee Arthroplasty: The WORQ. *The Journal of Arthroplasty*, **29**, 1169-1175.e2. <https://doi.org/10.1016/j.arth.2014.01.016>
- [13] Elangovan, N. and Sundaravel, E. (2021) Method of Preparing a Document for Survey Instrument Validation by Experts. *MethodsX*, **8**, Article ID: 101326. <https://doi.org/10.1016/j.mex.2021.101326>
- [14] Naderifar, M., Goli, H. and Ghaljaie, F. (2017) Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. *Strides in Development of Medical Education*, **14**, e67670. <https://doi.org/10.5812/sdme.67670>
- [15] First, M.B., et al. (2016) User's Guide for the SCID-5-CV Structured Clinical Interview for DSM-5® Disorders: Clinical Version. American Psychiatric Publishing, Inc.
- [16] Shankman, S.A., Funkhouser, C.J., Klein, D.N., Davila, J., Lerner, D. and Hee, D. (2017) Reliability and Validity of Severity Dimensions of Psychopathology Assessed Using the structured Clinical Interview for DSM-5 (SCID). *International Journal of Methods in Psychiatric Research*, **27**, e1590. <https://doi.org/10.1002/mpr.1590>

- [17] Kline, R.B. (1999) Book Review: Psychometric Theory (3rd Ed.). *Journal of Psychoeducational Assessment*, **17**, 275-280. <https://doi.org/10.1177/073428299901700307>
- [18] Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quinonez, H.R. and Young, S.L. (2018) Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, **6**, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>
- [19] STATA CORP (2021) Stata 17 Base Reference Manual. Stata Press.
- [20] Arbuckle, J. (2023) Amos 29.0 User's Guide. IBM SPSS.
- [21] Corporation, I. (2022) IBM SPSS Statistics Version 29. IBM.
- [22] Drost, E.A. (2011) Validity and Reliability in Social Science Research. *Education Research and Perspectives*, **38**, 105-123.
- [23] Bollen, K.A. (1989) Structural Equations with Latent Variables. Wiley. <https://doi.org/10.1002/9781118619179>
- [24] Ebadi, A., et al. (2017) Translation, Development and Psychometric Properties of Health Related Measures—Part 2: Construct Validity, Reliability and Responsiveness. *Payesh (Health Monitor)*, **16**, 445-455.
- [25] Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., et al. (2007) Quality Criteria Were Proposed for Measurement Properties of Health Status Questionnaires. *Journal of Clinical Epidemiology*, **60**, 34-42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- [26] Hutcheson, G. (1999) The Multivariate Social Scientist. SAGE Publications, Ltd. <https://doi.org/10.4135/9780857028075>
- [27] Cortina, J.M. (1993) What Is Coefficient Alpha? An Examination of Theory and Applications: Introductory Statistics Using Generalized Linear Models. *Journal of Applied Psychology*, **78**, 98-104. <https://doi.org/10.1037//0021-9010.78.1.98>
- [28] Weir, J.P. (2005) Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the Sem. *Journal of Strength and Conditioning Research*, **19**, 231-240. <https://doi.org/10.1519/00124278-200502000-00038>
- [29] Colton, D. and Covert, R.W. (2007) Designing and Constructing Instruments for Social Research and Evaluation. John Wiley & Sons.
- [30] de Vet, H.C.W., Terwee, C.B., Knol, D.L. and Bouter, L.M. (2006) When to Use Agreement versus Reliability Measures. *Journal of Clinical Epidemiology*, **59**, 1033-1039. <https://doi.org/10.1016/j.jclinepi.2005.10.015>
- [31] Terwee, C.B., Peipert, J.D., Chapman, R., Lai, J., Terluin, B., Cella, D., et al. (2021) Minimal Important Change (MIC): A Conceptual Clarification and Systematic Review of MIC Estimates of PROMIS Measures. *Quality of Life Research*, **30**, 2729-2754. <https://doi.org/10.1007/s11136-021-02925-y>
- [32] Koedel, C. and Betts, J. (2010) Value Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation. *Education Finance and Policy*, **5**, 54-81. <https://doi.org/10.1162/edfp.2009.5.1.5104>
- [33] Schluter, M.G., Hodgins, D.C., Stea, J.N. and Kilborn, M.L. (2022) Promoting Self-Change in Cannabis Use Disorder: Findings from a Randomized Trial. *Frontiers in Psychiatry*, **13**, Article 1015443. <https://doi.org/10.3389/fpsy.2022.1015443>
- [34] Baker, F.B. (2001) The Basics of Item Response Theory. ERIC.
- [35] Hu, L. and Bentler, P.M. (1999) Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**, 1-55.

<https://doi.org/10.1080/10705519909540118>

- [36] McDonald, R.P. (2013) *Test Theory: A Unified Treatment*. Psychology Press.
- [37] Masson, L., MCNeill, G., Tomany, J., Simpson, J., Peace, H., Wei, L., *et al.* (2003) Statistical Approaches for Assessing the Relative Validity of a Food-Frequency Questionnaire: Use of Correlation Coefficients and the Kappa Statistic. *Public Health Nutrition*, **6**, 313-321. <https://doi.org/10.1079/phn2002429>
- [38] Byrne, B.M., Shavelson, R.J. and Muthén, B. (1989) Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin*, **105**, 456-466. <https://doi.org/10.1037//0033-2909.105.3.456>

Appendix

تست تشخیص اختلال مصرف کانابیس - تجدید نظر شده نسخه ایرانی (CUDIT-R-Pr)

نام و نام خانوادگی (در صورت تمایل): سن: جنسیت: تأهل: تحصیلات: شغل: محل زندگی (در صورت تمایل):

آیا در شش ماه گذشته از ماری جوانا (گل یا حشیش) استفاده کرده‌اید؟ **بله/خیر**

اگر پاسخ بله است، لطفاً به سؤالات زیر در مورد مصرف ماری جوانا پاسخ دهید و در ستون پاسخ را علامت بزنید.

۱. چند وقت هست که از ماری جوانا استفاده می‌کنید؟

هرگز	۰	۱	۲	۳	۴
ماهانه یا کمتر	۴-۲	بار در ماه	۲-۳	بار در هفته	۴ بار یا بیشتر در هفته

۲. چند ساعت در یک روز معمولی که ماری جوانا مصرف می‌کردید، «دود» می‌کردید؟

کمتر از ۱	۰	۱	۲	۳	۴
۲-۱	۳ یا ۴	۵ یا ۶	۷ یا بیشتر		

۳. ۶ ماه گذشته چند بار متوجه شدید که نمی‌توانید مصرف ماری جوانا را متوقف کنید و به مصرف آن ادامه داده‌اید؟

هرگز	۰	۱	۲	۳	۴
کمتر از ماهانه	ماهانه	هفتگی	روزانه یا تقریباً روزانه		

۴. ۶ ماه گذشته چند بار نتوانستید کاری را که معمولاً از شما انتظار می‌رفت، به دلیل مصرف ماری جوانا انجام دهید؟

هرگز	۰	۱	۲	۳	۴
کمتر از ماهانه	ماهانه	هفتگی	روزانه یا تقریباً روزانه		

۵. ۶ ماه گذشته چند بار وقت زیادی را صرف تهیه، مصرف یا از بین بردن آثار بعد از مصرف ماری جوانا کرده‌اید؟

هرگز	۰	۱	۲	۳	۴
کمتر از ماهانه	ماهانه	هفتگی	روزانه یا تقریباً روزانه		

۶. در ۶ ماه گذشته، چند بار پس از مصرف ماری جوانا در حافظه یا تمرکز خود مشکل داشته‌اید؟

هرگز	۰	۱	۲	۳	۴
کمتر از ماهانه	ماهانه	هفتگی	روزانه یا تقریباً روزانه		

۷. چند بار از ماری جوانا در شرایطی استفاده کرده‌اید که از نظر فیزیکی می‌توانسته عواقب خطرناکی برای شما داشته باشد مانند رانندگی، استفاده از ماشین آلات، یا مراقبت از کودکان.

هرگز	۰	۱	۲	۳	۴
کمتر از ماهانه	ماهانه	هفتگی	روزانه یا تقریباً روزانه		

۸. آیا تا به حال به کاهش یا ترک مصرف ماری جوانا فکر کرده‌اید؟

هرگز	۰	۱	۲	۳	۴
بله، نه بیشتر از ۶ ماه	بله، در طول ۶ ماه				

• امتیازدهی سوال ۱ تا ۷ به صورت ۰ تا ۴ است، اما امتیازدهی سوال ۸ به صورت ۰، ۲، ۴ می‌باشد. نمره نهایی:

• استفاده از این پرسشنامه در صورت قید سازندگان و مترجمان آن بلامانع می‌باشد. این پرسشنامه توسط محمود امیری نیا و پریسا آقازاده در ایران (۱۴۰۰) از نسخه انگلیسی (Adamson, SJ, et al. (2010) ترجمه شده است و مورد اعتباریابی قرار گرفته است.

Figure A. Persian version of the Cannabis Use Disorder Identification Test—Revised (CUDIT-R-Pr).

The Cannabis Use Disorder Identification Test - Revised (CUDIT-R)

Have you used any cannabis over the past six months? Yes _____ No _____

If you answered "Yes" to the previous question, please answer the following questions about your cannabis use. Circle the response that is most correct for you in relation to your cannabis use over the *past six months*.

1. How often do you use cannabis?

Never 0	Monthly or less 1	2-4 times a month 2	2-3 times a week 3	4+ times a week 4
------------	----------------------	------------------------	-----------------------	----------------------

2. How many hours were you "stoned" on a typical day when you had been using cannabis?

Less than 1 0	1 or 2 1	3 or 4 2	5 or 6 3	7 or more 4
------------------	-------------	-------------	-------------	----------------

3. How often during the past 6 months did you find that you were not able to stop using cannabis once you had started?

Never 0	Less than monthly 1	Monthly 2	Weekly 3	Daily/almost daily 4
------------	------------------------	--------------	-------------	-------------------------

4. How often during the past 6 months did you fail to do what was normally expected from you because of using cannabis?

Never 0	Less than monthly 1	Monthly 2	Weekly 3	Daily or almost daily 4
------------	------------------------	--------------	-------------	----------------------------

5. How often in the past 6 months have you devoted a great deal of your time to getting, using, or recovering from cannabis?

Never 0	Less than monthly 1	Monthly 2	Weekly 3	Daily/almost daily 4
------------	------------------------	--------------	-------------	-------------------------

6. How often in the past 6 months have you had a problem with your memory or concentration after using cannabis?

Never 0	Less than monthly 1	Monthly 2	Weekly 3	Daily or almost daily 4
------------	------------------------	--------------	-------------	----------------------------

7. How often do you use cannabis in situations that could be physically hazardous, such as driving, operating machinery, or caring for children?

Never 0	Less than monthly 1	Monthly 2	Weekly 3	Daily/almost daily 4
------------	------------------------	--------------	-------------	-------------------------

8. Have you ever thought about cutting down, or stopping, your use of cannabis?

Never 0	Yes, but not in the past 6 months 2	Yes, during the past 6 months 4
------------	----------------------------------------	------------------------------------

This questionnaire was designed for self-administration and is scored by adding each of the 8 items:
 Question 1-7 are scored on a 0-4 scale
 Question 8 is scored 0,2, or 4

Score: _____

Scores of 8 or more indicate hazardous cannabis use, while scores of 12 or more indicate a possible cannabis use disorder for which further intervention may be required.

Adamson SJ, Kay-Lambkin FJ, Baker AL, Lewin TJ, Thornton L, Kelly BJ, and Sallman JD. (2010). An Improved Brief Measure of Cannabis Misuse: The Cannabis Use Disorders Identification Test - Revised (CUDIT-R). *Drug and Alcohol Dependence* 110:137-143.

Figure B. English version of the Cannabis Use Disorder Identification Test—Revised (CUDIT-R).