*Document Version*
Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](Link to publication record in King's Research Portal)

*Citation for published version (APA):*
Mihretu, A., Fekadu, A., Norton, S., Habtamu, K., & Teferra, S. (2022). Validation of the Problematic Khat Use Screening Test: A Cross-Sectional Study. *European Addiction Research*, *28*(4), 275-286. https://doi.org/10.1159/000522618

**Research Article**

# Validation of the Problematic Khat Use Screening Test: A Cross-Sectional Study

Awoke Mihretu[a]    Abebaw Fekadu[a, b, c]    Sam Norton[d]    Kassahun Habtamu[e]
Solomon Teferra[a, f]

[a]Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia; [b]Centre for Innovative Drug Development and Therapeutic Trials for Africa (CDT-Africa), College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia; [c]Global Health & Infection Department, Brighton and Sussex Medical School, Brighton, UK; [d]Health Psychology, Institute of Psychiatry Psychology & Neuroscience, King's College London, London, UK; [e]School of Psychology, College of Education and Behavioral Studies, Addis Ababa University, Addis Ababa, Ethiopia; [f]Harvard T.H. Chan School of Public Health, Boston, MA, USA

**Abstract**

*Aim:* The study aimed to evaluate the psychometric properties of the Problematic Khat Use Screening Test (PKUST-17) in Ethiopia. *Methods:* A validation study of PKUST-17 was carried out among 510 khat users, using a house-to-house survey. Confirmatory factor analysis and 2-parametric item response theory (IRT) were used to evaluate the construct validity of PKUST-17. We also used Spearman's rank-order correlation coefficient and other test statistics to assess the convergent validity of PKUST-17 with depression symptoms, functional impairment, and other characteristics of participants. We generated latent classes of problematic khat use using latent profile analysis (LPA) and validated the classes using multinomial logistic regression. *Results:* The data confirm the unidimensional model of the PKUST-17. The internal consistency of PKUST-17 was excellent (Cronbach's alpha = 0.93). IRT discrimination parameters indicated that each item had a strong ability to distinguish participants across the spectrum of problematic khat use ($\alpha$ thresholds range from 1.02 to 2.9). The items were fairly or moderately severe to be endorsed by participants ($\beta$ thresholds vary from 1.43 to 5.57). The LPA identified three latent classes which have severity differences: mild (34%), moderate (34%), and severe (32%) problematic khat use. Depression symptoms, functional impairment, and other khat use patterns were also associated with moderate and severe problematic khat use class membership compared to mild problematic khat use class. *Conclusion:* We found that the PKUST-17 is a culturally appropriate, brief, easy to use, and psychometrically sound screening test. PKUST-17 can be used to screen khat users with different levels of risk for providing stepped care at different healthcare levels, including integration of services in primary care. Future studies need to test the predictive capacity of the PKUST-17 for khat-related harms.

© 2022 The Author(s).
Published by S. Karger AG, Basel

## Introduction

There has been a long history of chewing khat (*Catha edulis* [Vahl] Forssk. ex Endl), an evergreen shrub/tree which contains an amphetamine-like stimulant [1]. Khat is a type of natural amphetamine because its ingredients, such as cathinone, cathine, and norephedrine, have a sim-

Correspondence to:
Sam Norton , sam.norton@kcl.ac.uk

ilar chemical structure to amphetamine [2]. United Nations Convention on Psychotropic Substances scheduled cathinone and cathine in schedule I and IV, respectively, but the plant khat itself is not scheduled [3]. The khat plant's fresh leaves typically include 114, 83, and 44 mg of cathinone, cathine, and norephedrine per 100 g, respectively [4].

Khat use is common in many East African countries and the Arabian Peninsula, and also East African immigrants who live in Western countries. The current prevalence of khat use in adults is estimated to be as high as 67.9% in Yemen, 59% in Somalia [5, 6], and 15.3% in Ethiopia [7]. A systematic review found that the prevalence of khat use among adolescents, specifically high school and higher education students, is 16.7% in Ethiopia [8].

Khat is used for various sociocultural reasons in different settings [9, 10]. For example, people often report that they chew khat to stay alert during praying or studying religious issues and to connect with people for social gatherings during weddings and funeral ceremonies.

Regarding mental health and khat use, psychotic symptoms have been reported among heavy khat users [11]. However, a systematic review concluded that there is no evidence for an association between khat use and severe mental disorders [12], but the impact of khat use on common mental disorders could be significant [13].

Although little is known about which pattern of khat use is associated with adverse consequences, problematic khat use patterns might be associated with different adverse effects than khat use per se. Thus, recently, problematic khat use has been a concern of researchers and policymakers [14–16]. We have done a series of studies focusing on what constitutes problematic khat use [17, 18]. We found problematic khat use is a dysfunction of khat use characterized by frequent use, chewing for long hours, khat use-specific and khat use-related financial harms, and different withdrawal experiences [18].

While there are many screening tools for other psychoactive substances such as problematic cannabis and alcohol use [19, 20], little progress had been made to measure problematic khat use. Diagnostic and Statistical Manual (DSM-5) criteria for stimulant use disorders were found to be valid for problematic khat use in Ethiopia [21], but the validation was limited to construct validity and therefore there is no strong psychometric evidence against DSM-5 to be used among the general population in Ethiopia. DSM-5 has also limited utility for screening problematic khat use by nonmental health professionals since it was designed for diagnosis in clinical settings. The DSM-5 is not designed for lay provider use

and would require an intensive training to enable a lay provider to use it appropriately. Thus, there is a need for a khat screening measure that is more easily administered by lay providers.

The severity of dependence scale [22] is currently the most widely used measure of problematic khat use, but it only measures a narrow concept of problematic khat use. Thus, taking previous theoretical and methodological lessons, we have developed the Problematic Khat Use Screening Test (PKUST-17) following standardized and rigorous procedures [17, 18]. PKUST-17 is a problematic khat use screening test with 17 items focusing on the frequency of khat use, amount of time spent chewing khat, financial problems, and different withdrawal experiences [23]. The current study aimed to evaluate the psychometric properties (internal consistency and concurrent, convergent, and construct validity) of PKUST-17 in the Gurage Community, South-central Ethiopia.

## Materials and Methods

*Study Design*
We used a community-based cross-sectional survey to investigate the psychometric properties of the PKUST-17.

*Study Setting*
The study was conducted in Wolkite town and Kebena district, Gurage zone, located 158 km south of Addis Ababa, the capital city of Ethiopia. The area comprises both urban and rural residents. We have reported detailed descriptions of the study setting in our previous formative qualitative study [17].

*Source Population, Sample Size, and Sampling*
This validation study's target population was all adults (18 years and above) who have lived in Wolkite town and Kebena district for at least 6 months. All the five accessible rural subdistricts of the Kebena district, surrounding Wolkite town, and three subdistricts from Wolkite town were selected and included in the study. Subdistrict, kebele, is the smallest administrative unit in the study setting. We found a sampling frame from health posts in each subdistrict. We used a two-stage random selection method called the Kish method [24]. We first randomly selected the households, and then we chose participants among the eligible persons within a household. Informed by a pilot study and rules of thumb recommended in the literature, we used a sample size of 30 participants per item for confirmatory factor analysis (CFA) and latent profile analysis (LPA) [25, 26]. Thus, the required sample size was 510 given the 17 items are included. Simulation studies for latent class and LPA suggested sample sizes of 300–500 are typically appropriate so our sample is also sufficient for this type of analysis [27]. Since we found it was not feasible to interview all samples by clinicians using DSM-5 criteria, convergent validity evaluation, we did subsampling. Thus, using a statistical formula to determine how many of the khat users should be interviewed for DSM-5 [28], we determined and randomly selected a sample of 232 participants.

## Measures

### Sociodemographic Characteristics and Patterns of Khat Use

We used a structured questionnaire to collect data on sociodemographic characteristics (sex, age, marital status, relative wealth, residence, educational, and academic status) and patterns of khat use (amount of khat use, frequency of khat use, duration of khat use, and time of khat session).

### Problematic Khat Use Screening Test

PKUST-17 is used to measure problematic khat use. PKUST-17 is a newly developed screening tool with 17 items and a 5-point (0–4) Likert scale response format. The total scores of the tool range from 0 to 68. PKUST-17 was developed using a series of studies; systematic review, qualitative study, experts' consensus meetings, cognitive interviewing, and pilot study [17, 18]. The systematic review and qualitative study aimed to conceptualize the construct problematic khat use and to develop a pool of items that constitute problematic khat use. Other studies were used for item refinement and item reduction. We reported the development and initial psychometric properties of the PKUST-17 in another study [23].

### DSM-5 Criteria for Stimulant Use Disorders

We used the DSM-5 criteria of stimulant use disorders [29] to diagnose people with khat use for khat use disorder using mental health professionals. Stimulant use disorder criteria of the DSM-5 have 11 items in four domains: impaired control, pharmacological criteria, social impairment, and risky use. A previous study in Ethiopia suggested DSM-5 has an acceptable construct validity to measure khat use disorder [30]. In the current sample, the internal consistency (Cronbach's alpha) of the FDM-5 was 0.95.

### Depression

The nine items Patient Health Questionnaire (PHQ-9) was used to screen depression symptoms in the general population [31]. PHQ-9 has been validated in Ethiopia, both in rural and urban settings [32, 33]. They found PHQ-9 has acceptable psychometric properties. It is also a one-factor measure with a cut-off point of five and above in rural settings and ten and above in urban settings. In the urban setting, PHQ-9 sensitivity was 86% and specificity was 67% [33]. In the rural setting PHQ-9 sensitivity was 83.5% and specificity was 74.7% [32]. In the current sample, the internal consistency (Cronbach's alpha) of this questionnaire was 0.76.

### Disability

We used the World Health Organization Disability Assessment Schedule (WHODAS 2.0) to measure disability [34]. WHODAS 2.0 has 12 items and measures cognition (understanding and communicating), community participation, life activities (home, academic, and occupational functioning), self-care, getting along with people, and mobility (getting around) [34, 35]. The instrument was validated in Ethiopia and reported acceptable psychometric properties [35]. WHODAS 2.0 has also been recommended to be an essential patient-reported outcome measure for all DSM-5 or ICD-11 substance use or psychiatric disorder [36]. In the current sample, the internal consistency (Cronbach's alpha) of WHODAS 2.0 was 0.89.

### Social Support

The three items Oslo social support scale (Oslo-3) was used to measure social support. Oslo-3 measures perceived social support level and have overall scores ranging from 3 to 14. Lower values indicate poor social support. Oslo-3 has also been used previously in Ethiopia without any indication of issues with validity in this population [37, 38]. In the current sample, the internal consistency (Cronbach's alpha) of Oslo-3 was 0.81.

### Stressful Life Events

A list of threatening events (LTE) is a list of significant and threatening events such as loss of relationships, death of close persons, and jail [39]. LTE items are dichotomous with the "No" or "Yes" response format. LTE has good reliability (test-retest reliability of 0.61–0.87) and validity (convergent and construct validity) [39]. LTE questionnaire has also been previously used in rural Ethiopia again without any indication of validity issues [38].

### Alcohol Use Disorder

We used the Alcohol Use Disorder Identification Test (AUDIT), developed by WHO, to measure alcohol use disorder [40]. It has ten items assessing alcohol consumption behavior (amount of alcohol, frequency of drinking, and adverse consequences related to alcohol use) in the past 12 months. Items in AUDIT have polytomous response formats ranging from 0 to 4; their total score ranges from 0 to 40. The cut-off for problematic alcohol use is eight or more points [41]. In Ethiopia, AUDIT has been used in several studies, and its internal consistency was found to be very high (Cronbach's alpha = 0.84) [42]. In the current sample, the internal consistency (Cronbach's alpha) of AUDIT-10 was 0.8.

### Household Food Insecurity

We used Household Food Insecurity Access Scale (HFIAS) to measure household food insecurity. It has nine items with a polytomous response format, ranging from 1 to 3 categories [43]. It measures anxiety or uncertainty about food supply, insufficient quality of food both in variety and preference, and inadequate quantity of food supply. HFIAS has been used and validated in Ethiopia before, both in rural and urban settings, and found to have acceptable psychometric properties [44]. In the current sample, internal consistency (Cronbach's alpha) of HFIAS was 0.81.

### Data Collection Procedure

Trained lay data collectors interviewed participants using the measures mentioned above. Masters mental health clinicians examined khat users to diagnose problematic khat use using DSM-5. Khat users were interviewed for both PKUST-17 and DSM-5, but the sequence of interviews for both tools was random. Either lay data collectors go first or clinicians. This was intended to reduce any bias introduced by administering one tool before the other [45].

### Data Analysis

Convergent validity was assessed using Spearman's Rho correlation coefficient to calculate the association between total scores of problematic khat use screening scale and DSM-5 criteria, the HIFAS, WHODAS 2.0, depression/PHQ-9, social support, and Stressful Life Events/LTE scales. We used nonparametric statistics, Kruskal-Wallis, and Mann-Whitney U test, to evaluate the difference of PKUST-17 scores across different patterns of khat use and other characteristics of participants. We also used Cronbach's alpha to assess internal consistency [46].

CFA was used to compare the fit of the data with the unidimensional PKU theoretical model and evaluate the items' ability. Indices of acceptable fit for CFA include; Root Mean Square Error of Approximation close to 0.06, Standardized Root Mean Residual close to 0.06, and Comparative Fit Index close to 0.95 [47]. Item Response Theory (IRT) models, specifically graded response models, were then used to determine item functioning in terms of item response difficulty and discrimination. Under this model, a separate difficult parameter is estimated for each response category for an item and represents the level of the latent trait at which 50% of the samples are expected to endorse the response category. One discrimination parameter, which is related to the concept of the factor loading in CFA, is estimated per item indicates the degree to which an item can differentiate between different levels of the latent trait. These parameters are used to graph the item characteristic curve, which indicates the expected probability of responding to each item category across the range of the latent trait. We assessed the IRT assumptions: unidimensionality, local independence, and monotonicity. The residual correlation matrix of the unidimensional CFA with a value of 0.2 above the average residual correlation was considered a critical value to violate the assumption of local independence [48]. Item discrimination parameter greater than 4 ($\alpha < 4$ for all items in the current study) provides evidence that there are no items with local dependence. The item characteristic curve's shape looks like the probability of endorsing an item is not decreasing; thus, it is possible to declare the monotonicity of the PKUST-17 items [49].

A logistic regression-based method, which was preferred over Mantel-Haenszel-based techniques, was used for IRT differential item functioning (DIF) evaluation [50]. Due to the presence of a low sample size per response category, the data could not be fitted with an ordinal logistic regression modeling technique to assess test items for DIF. Thus, polytomous items and discrete exposure variables were recoded to dichotomous variables (response category "0" and "1" vs. "2," "3," "4").

LPA was used to uncover latent classes, typologies, of problematic khat use. LPA, like latent class analysis, is a technique used for discovering latent groups in data by obtaining the probability that individuals belong to distinct groups [51]. Both LPA and latent class analysis are model-based methods for estimating population characteristics and adjusting for measurement error. They also use probabilities as the basis for an interpretation of statistical outputs and flexible treatment of variance among classes [52, 53]. Besides, they have also clinical implications such as designing common interventions based on latent classes' shared characteristics [54, 55]. In the current study, LPA was applied to uncover latent classes from the continuous PKUST-17 scores. The optimal number of classes to extract was determined using standard model fit statistics. Although there is no consensus across the literature about the absolute criteria for latent class determination, the following fit statistics are recommended. (a) Bayesian information criterion (BIC) and sample size adjusted BIC [53, 56], (b) Akaike information criteria (AIC), (c) Likelihood tests (i.e., Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test) [57]. The Likelihood ratio test provides a $p$ value, which indicates if one model is statistically better than another [53].

Lower BIC and AIC indicate a better fit of the classes. In addition to these parameters, we also report cs, such as Entropy [58], which indicates the accuracy with which the model defines classes. This is useful since LPA models are probabilistic with each person given a probability of being included in each class. Values closer to one indicate greater certainty in class assignment [59].

The number of classes to extract was based partly on model fit statistics but also on the theoretical interpretability of the different classes [53, 60]. After considering the above criteria, study reporting was guided by protocols of reporting latent class models [61]. Descriptive statistics and test statistics (Kruskal-Wallis and $\chi^2$ test) were used to summarize the data and examine differences among 3-class participants. Multinomial logistic regression was used to validate the three latent classes of PKU. We used Latent GOLD 5.1 [62], STATA 16 [63], and SPSS 23 AMOS [64] computer software packages for data analysis.

*Data Quality Assurance*

Those involved with data collection were selected based on their experience of administering DSM-5 in their routine clinical practice and many years' experience using the instruments involved in this study. Then, after brief orientation, we checked inter-rater reliability and found a very high level of agreement. The Kappa was 0.9 for clinical diagnosis. We trained lay interviewers for 1 day in the use of an earlier version of the instrument used in this study and have experience administering the current screening tool in a series of three studies as part of the development and validation of the instrument; pretest, bigger pilot study, and the current validation study. Since the current tool is well structured, we found that inter-rater reliability was not a problem. The measures also appear good in layout, and due attention was given to coding items and responses. The PI did close supervision and a daily check of the data for completeness and other data collection gaps during fieldwork. The sequence of administering DSM-5 and PKUST-17 was random to avoid potential biases. Both lay interviewers and clinicians were masked about the result for the problematic khat use status of the participant interviewed for the other tool (PKUST-17 or DSM-5). Data entry with consistency check was conducted using EpiData [65].

*Ethical Considerations*

This study was approved by the Institutional Review Board of the College of Health Sciences (ref 008/18/psy), Addis Ababa University. The information sheet was prepared and presented to each participant to allow free and informed decision to participate in the study. The study maintained privacy and confidentiality during data collection, handling, and reporting. All standard national guidelines for COVID-19 precautions, including wearing a facemask, physical distancing, and taking hand hygiene measures, were maintained.

## Results

*Sociodemographic Characteristics of Participants*

In total, 506 people participated in this validation study. The mean (±SD) age of the participants was 34.4 (±13) (online suppl. Additional File 1; online suppl. Table 1; for all online suppl. material, see www.karger.com/doi/10.1159/000522618). About 41% ($n = 208$) were from rural area. One hundred and fifteen (24%) participants

had no formal education. Only very few participants endorse high relative wealth. 42% of the participants rated their wealth as medium or high relative to other people they know in their area.

*Descriptive Statistics of the PKUST-17 Items*

The scale mean (±SD) was 19.6 (±14.5). Internal consistency (Cronbach's alpha) was 0.93. Among the 17 items, the highest mean was for the item about the frequency of khat use, and the lowest mean was for the item about the increased amount of khat over time. There were no floor and ceiling effects. For the overall scale, the proportion of participants with ceiling and floor scores were 1% and 0.2%, respectively. The highest mean score (2.92) was for the item about frequency of khat use and the lowest mean score (0.62) was for the item about increased about of khat over time (online suppl. Additional File 1; online suppl. Table 2).

*Convergent Validity*

The Spearman rank-order correlation between PKUST-17 scores and DSM-5, WHODAS, Oslo-3, PHQ-9, and LTE were statistically significant. There was a moderate correlation between PKUST-17 scores and DSM-5 ($r = 0.57$; $p < 0.05$) and PHQ-9 scores ($r = 0.5$; $p < 0.05$). The correlation between PKUST-17 scores and LTE ($r = 0.24$; $p < 0.05$), HIFAS ($r = 0.22$; $p < 0.05$), Oslo-3 ($r = -0.3$; $p < 0.05$), and WHODAS scores ($r = 0.33$; $p < 0.05$) was weak. The correlation between PKUST-17 scores and social support (Oslo-3) was negative as hypothesized.

*Psychosocial Variables, Patterns of Khat Use, A Summary of PKUST-17 Score*

Mann-Whitney U test indicated a statistically significant difference in PKUST-17 scores across gender, reasons for khat use, age of onset of khat use, chewing khat daily or not, chewing khat in the morning or not, level of stress, depression symptoms, and food insecurity status. Higher PKUST-17 scores were observed among males, chewing for relieving distress, started chewing khat before 18 years, with more stressful experiences and depression symptoms. PKUST-17 scores did not significantly vary in residence, relative wealth or income status, amount of khat per session, whether they chew alone or with others, and physical health condition status. The Dunn post hoc test for Kruskal-Wallis found that participants who chew in their homes scored lower in PKUST-17 than others who chew at other places. It also indicated that married participants also scored higher in PKUST-17 scores than never-married participants. Dunn post hoc test par-

ticipants who were employed also recorded more on PKUST-17 than participants who were farmers or housewives in their occupation (online suppl. Additional File 1; online suppl. Table 3).

*Confirmatory Factor Analysis and Item Response Theory*

The items were considered to be suitable for CFA since the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.95, and Bartlett's test of sphericity was significant ($\chi^2 = 4,769.9$, df = 136, $p < 0.05$). The data confirmed a unidimensional model of problematic khat use. All 17 items loaded onto the resulting factor with an item factor loading of 0.48 or above except one item about the amount of time spent while chewing khat (PK16-item) with a factor loading of 0.35 (online suppl. Additional File; online suppl. Table 4).

IRT discrimination (slope) parameters indicate the items' ability to distinguish participants across the spectrum of problematic khat use ranged from 1.02 (increased amount over time) to 2.9 (depressed mood when not chewing) except one item about the amount of time spent chewing khat with value of 0.75. All of the item response categories had difficulty estimates above one (range 1.43–5.57) except for one item about khat use frequency. Taken together, this indicates that the scale distinguishes effectively between those with different levels of problematic khat use risk, except for those at the low end of the trait with a low likelihood of problematic khat use. According to the IRT test information curve, the current problematic khat use screening tool provides a lot of information (i.e., high reliability) in the moderate range of the latent trait (problematic khat use) (Fig. 1–3).

Both nonuniform and uniform DIF was found for residence, sex, and education. Three items were flagged for DIF regarding residence, sex, and education. The uniform and nonuniform DIF findings are presented in an additional file (online suppl. Additional File 2).

*Latent Profile Analysis*

LPA models extracting 1 through 4 classes were considered. The 2-class model provided the best fit to the data according to BIC and the 4-class model by the AIC and SABIC. However, the 4-class model included a class that contained very few individuals that was not informative theoretically. Instead, we chose the 3-class model as this provided the most useful interpretation of the data with information and had BIC, AIC, and SABIC values very similar to those of the other models (online suppl. Additional File 1; online suppl. Table 5).
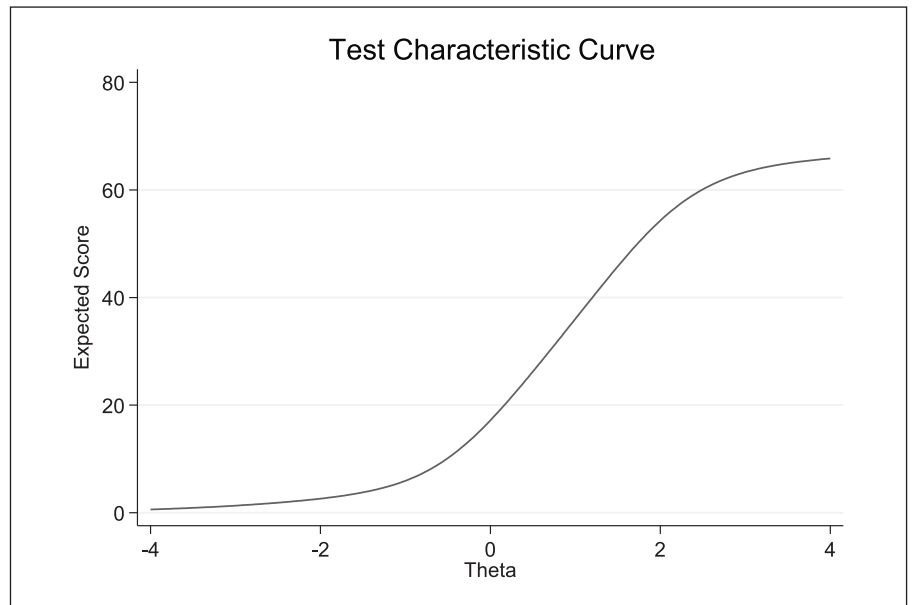
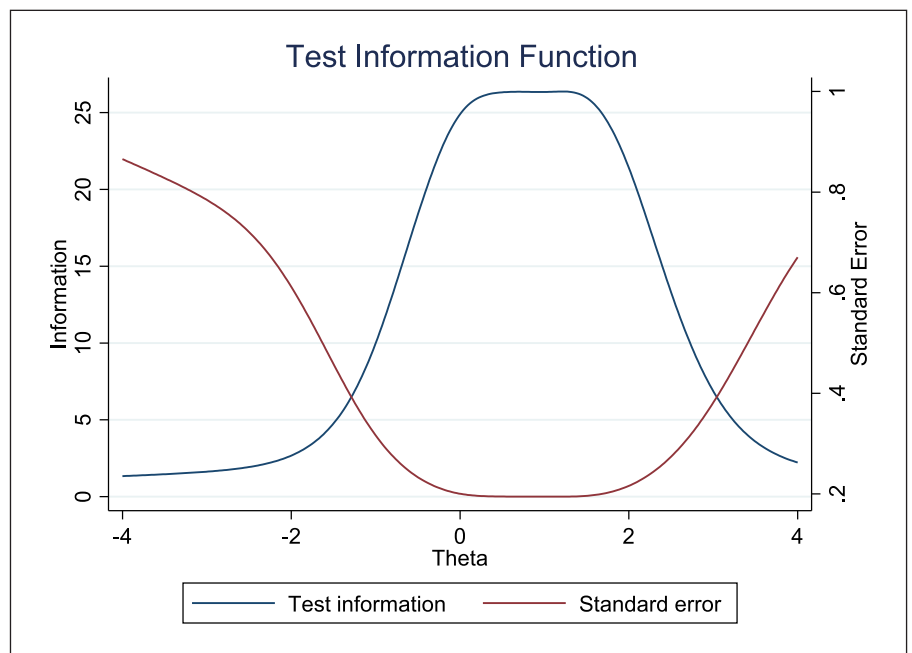**Fig. 1.** IRT test characteristics graph for PKUST-17.



**Fig. 2.** IRT test information function and standard error graph.

*Endorsement Probability of the PKUST-17 Items*

Class 1 participants' endorsement probabilities were below 50% for many items except for an item about the frequency of khat use (PK1). Class 2 participants' endorsement probabilities for PKUST-17 items were above 50% except for an item about an increased amount of khat overtime (PK2). Thus, there is more severity and some other profile difference among the three classes. Thus, we named class 1 = "mild problematic khat users," class 2 =

"moderate problematic khat users," and class 3 =" severe problematic khat users."

*The Profile of the Three Problematic Khat Use Classes*

The mean PHQ-9, WHODAS, and HIFAs scores for severe problematic khat users were 7.3, 9.13, and 9, respectively. In comparison, the mean PHQ-9, WHODAS, and HIFAs scores for mild problematic khat users were 2.2, 3.5, and 5.1, respectively. Kruskal-Wallis test found a sig-
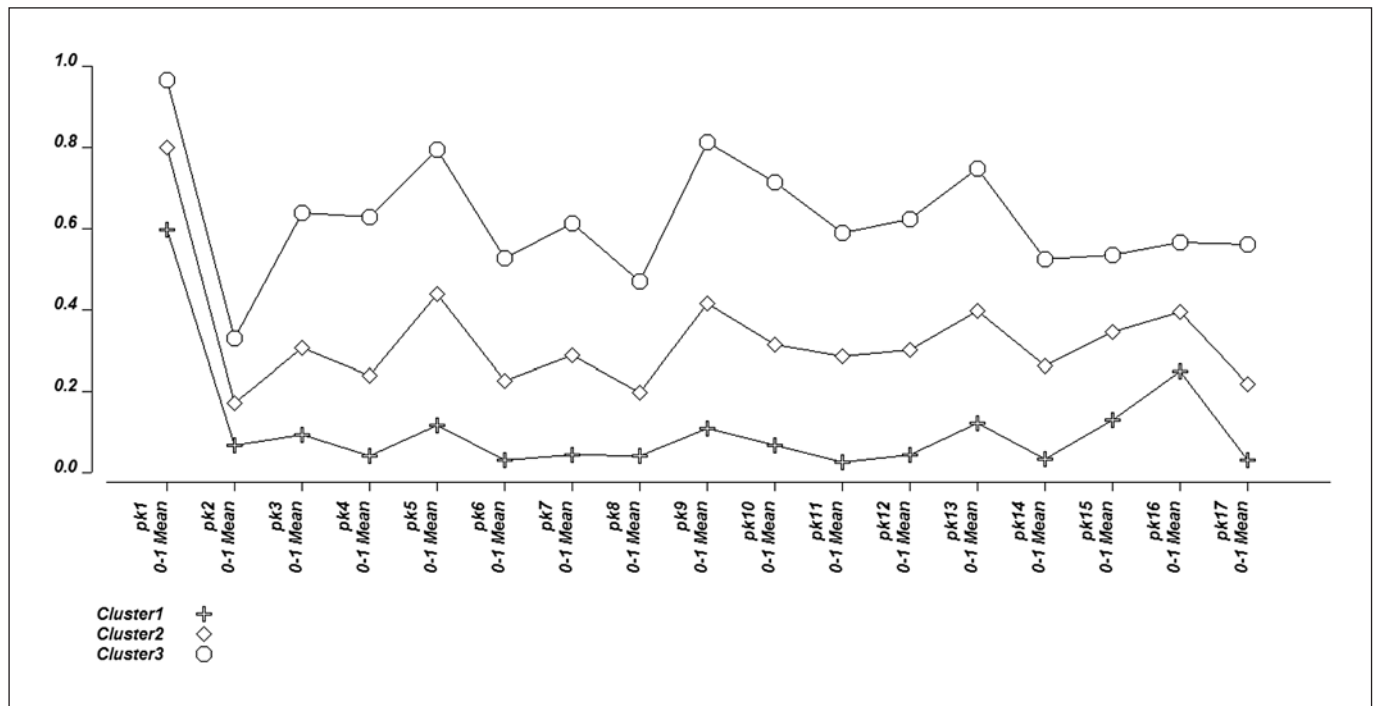
Mihretu/Fekadu/Norton/Habtamu/
Teferra

**Fig. 3.** Endorsement probabilities of PKUST-17 items per three latent problematic khat use classes.

**Table 1.** Multinomial regression model of problematic khat use classes predicted by different variables

| Covariates | Sever versus mild AOR (95% CI) | Moderate versus mild AOR (95% CI) |
|---|---|---|
| WHODAS scores | **1.16 (1.09–1.25)** | **1.12 (1.07–1.16)** |
| PHQ-9 scores | **1.17 (1.06–1.3)** | **1.18 (1.09–1.29)** |
| HIFAS scores | 0.04 (0.97–1.11) | 0.99 (0.94–1.04) |
| Oslo-3 scores | **0.76 (0.64–0.91)** | 0.93 (0.82–1.05) |
| Any chronic health problem | 1 (0.45–2.2) | 1.44 (0.8–2.58) |
| Two and more stressful life events | 0.88 (0.38–2.02) | 1.28 (0.72–2.28) |
| Regular morning khat use | **2.11 (1.04–4.3)** | **2.06 (1.02–4.17)** |
| Place of khat use | | |
|     Work place | 1.53 (0.61–3.8) | 1.83 (0.95–3.5) |
|     Khat café | 0.59 (0.14–2.48) | 0.99 (0.35–2.8) |
|     Chewing elsewhere | 1.93 (0.71–5.23) | **3.52 (1.78–6.94)** |
|     Reasons of khat use (social and functional reasons) | 0.64 (0.23–1.92) | 2.19 (0.73–6.53) |
|     Marriage (married) | 0.51 (0.2–1.3) | 0.85 (0.44–1.63) |
|     Sex (female) | 0.16 (0.73–6.66) | 5.31 (1.07–26.28) |
|     Age in years | 0.99 (0.96–1.02) | 0.94 (0.89–0.98) |
| Education | | |
|     Primary | 0.52 (0.19–1.41) | 0.72 (0.34–1.5) |
|     Secondary | 0.85 (0.29–2.47) | 0.49 (0.23–1.06) |
|     College and above | 0.57 (0.18–2) | 0.7 (0.31–1.57) |
| Amount of khat (USD 1 and above) | **2.23 (1.08–4.62)** | 1.13 (0.67–1.92) |

nificant difference in age, HIFAS, PHQ-9, and WHODAS scores across the three classes of problematic khat use. The Dunn post hoc test found that participants with moderate problematic khat use scored more on PHQ-9 and WHODAS than mild problematic khat users. Severe problematic khat users were also higher in their mean PHQ-9 and WHODAS scores than participants in the mild problematic khat use. Severe problematic khat users were older compared to mild and moderate problematic khat users. Moderate and Severe problematic khat users were also more food unsecured than mild problematic khat users (online suppl. Additional File 1; online suppl. Table 6).

*Latent Class Regression*

The likelihood of the participants' membership to moderate and severe problematic khat use classes is compared to mild problematic khat use (reference class) using multinomial logistic regression (Table 1: "Normal table"). Participants who regularly chew khat in the morning compared to those who did not chew had about 2.06 increase odds for being in the moderate problematic khat use class (AOR = 2.06; 95% CI [1.02–4.17]) and 2.11 increase odds for being in the severe problematic khat use class (AOR = 2.11; 95% CI [1.04–4.3]). A single score increase in depression symptoms (PHQ-9) had also 1.18 time more or 18% increase likelihood of being in the moderate and 1.17% or 17% increase likelihood of being in the severe problematic khat use class compared to mild problematic khat use class (AOR = 1.18; 95% CI [1.09–1.29]) and (AOR = 1.17; 95% CI [1.06–1.3]). Regarding functional impairment, a single score increase in WHODAS scores also had about 1.12 times more likely or 12% increase in the likelihood of being in the moderate and 1.16 times more likely or 16% increase likelihood of being in severe problematic khat use class compared to mild problematic khat use class (AOR = 1.12; 95% CI [1.07–1.16]) and (AOR = 1.16; 95% CI [1.09–1.25]). The odds of being in the moderate problematic khat use compared to mild problematic khat use for participants who chew khat elsewhere, including on the street relative to those who chew in their homes, were 3.52 (AOR = 3.52; 95% CI [1.78–6.94]). Table 1 shows the full and final multinomial logistic regression model of the three problematic khat use classes as predicted by exposure variables.

## Discussion

The current study found that PKUST-17 was a valid and potentially useful instrument among the general population, which adds to the existing literature that supports the use of valid and brief screening tools for identifying problematic substance use [41, 66, 67]. The strong correlation between PKUST-17 and DSM-5 criteria for stimulant use disorders indicates that both tools measure the same construct or very similar construct. Although there are similarities among the tools, the response categories were different. Polytomous items had more advantages to measure the severity and gain more information on the cost of their complexity for some people [32]. We excluded DSM-5 items such as impairment of control, a failure to fulfill major role obligations, and other items about social impairment from the item pool during the PKUST-17 development study [23]. The normative context of khat use in the study setting might be the reason to reduce the social impacts due to problematic khat use [17].

The IRT test information function graph indicated that the PKUST-17 was precise for people with moderate level of problematic khat use. Less information was provided for individuals with latent trait estimates below -3 or above 3. The current TIF graph has a bell shape, unlike diagnostic tests. The DSM-5 criteria for alcohol use disorder and peaked shape at the higher end of the continuum tap the cases more severely [68]. Thus, PKUST-17 had an excellent utility for the general population in the primary health care setting than in the clinical settings like diagnostic tests.

The LPA was able to identify three latent classes of problematic khat use (mild, moderate, and severe). The latent classes had more severity than typological differences consistent with the latent class of problematic alcohol and cannabis users and behavioral addiction [69–74]. The current prevalence of mild, moderate, and severe problematic khat use was 34%, 34%, and 32%, respectively, using PKUST-17. A previous study had reported 10.5%, 8.8%, and 54.5% prevalence of mild, moderate, and severe levels of problematic khat use using DSM-5 among the general population and university students in Ethiopia [21]. The current study is also similar to this previous study which found a 3-class model of problematic khat use using DSM-5 with 35%, 33%, and 32% proportion of participants [21]. In line with the previous study, the current study supports that a data-driven approach using latent models such as LPA is vital to uncover latent subgroups.

The findings from the current study using LPA also support the hypothesized relationship of severity of problematic khat use with psychosocial problems, khat use patterns, and functioning problems. Participants with more depression symptoms, functional impairment, and

poor social support were in severe or moderate problematic khat users' class. There was also evidence that stimulants like problematic khat use had been associated with depression [75, 76]. Moderate and severe problematic khat use could precipitate depression indirectly by impairing psychological adjustment and social support. Similar evidence has been reported that problematic khat users report more common mental health or depression symptoms and functional impairment or poor quality of life [21]. Depression symptoms and functioning problems did not significantly discriminate moderate and severe problematic khat use classes because it suggested further investigation about the two classes' distinctive nature using prospective studies. Patterns of khat use (chewing in the morning, increased amount of khat, and place of khat session) were also significantly associated with moderate or severe problematic khat use class membership compared to mild problematic khat user class. This resembles the existing knowledge of latent class membership of problematic alcohol [69] and cannabis use [77]. One previous study also found that patterns of use such as frequency and chewing khat in the morning were essential indicators of problematic khat use [21]. Generally, the LPA findings further infer that PKUST-17 could discriminate different types of PKU groups.

There are several strengths and limitations of the current study. Among the strengths, we attempted to overcome many of the previous studies' limitations that focused on measuring problematic khat use with the deductive (etic) approach only. Most extant instruments were also methodologically constructed using classical test theory with its psychometric limitations that directly impact severity measurement. We applied IRT and a data-driven method of validation, LPA. The cross-sectional nature of the current study did not examine the predictive validity of PKUST-17. Thus, the study would not indicate the responsiveness that could change the current screening tool's ability. The study lacked evidence to infer the application of PKUST-17 among adolescents with problematic khat use. Criterion validity could be the strongest but hard to achieve scale validity. We could not get a "gold" standard to evaluate PKUST-17 criterion validity. Since we hypothesis the current PKUST-17 is expansive and had many items which did not find in the DSM-5, we hesitated to consider DSM-5 as a real gold standard. Recoding variables to a different type than the original measurement might force the statistical model fitness during IRT DIF analysis. Thus, the lack of sufficient samples per response category to fit an ordinal logistic regression modeling technique and assess the test items for DIF was

also a limitation. Since khat is culturally a less acceptable behavior for women than men in the study setting in specific, and in Ethiopia in general, they less may be less likely to endorse use. Although the current study as well as the previous studies used random sampling, the prevalence of khat use among women is very low [78, 79]. Thus, much of the evidence about khat use, including the current findings mainly apply for men. The current measure does not introduce disparity between men and women.

## Conclusion

The PKUST-17 was the first culturally appropriate and psychometrically sound screening test that followed a rigorous methodology for scale development and validation. It would be useful for screening problematic khat users at different risk levels. The data confirmed the unidimensional model of problematic khat use. PKUST-17 had a moderate level of correlation with DSM-5 and PHQ-9 scores. The PKUST-17 scores were also different across variables such as reasons for khat use, household food insecurity access, daily use, occupation, chewing khat in the morning, place of khat use, and alcohol drinking. Thus, there was evidence for the convergent validity of the PKUST-17. IRT also indicated acceptable severity and discrimination for all PKUST-17 items. According to the IRT test information curve, the current problematic khat use screening tool provided high information, which is high reliability, in the moderate range of the latent trait (problematic khat use). The LPAs indicated a 3-class solution as the best-fitting model. The three distinct latent classes were different based on severity than typology. The numbers of participants for every three classes were proportional. Depression, social and functional impairments, and other typical indicators of problematic khat use such as chewing khat in the morning and amount of khat were statistically significant with problematic khat use class membership which further inferred evidence of validity for the construct problematic khat use and the discriminate ability of PKUST-17. In addition to acceptable psychometric properties, the current tool also conforms to the criteria for the relevance of a given test, such as simplicity, ease to understand, brief (maximum 20) items, capable of being self-administered requiring specific training for administration [80]. Further studies need to test the predictive capacity of the PKUST-17 for khat-related harms.

## Acknowledgments

## Statement of Ethics

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. We obtain Written informed consent was obtained from the participants. Ethical approval was secured from the Institutional Review Board of the College of Health Sciences, Addis Ababa University under protocol number 008/18/psy.

## Conflict of Interest Statement

The authors declare that they have no competing interests.

## Funding Sources

## Author Contributions

The first and the last authors A.M. and S.T. designed the protocol. A.F., S.N., K.H., and S.T. contributed on data analysis and manuscript write-up.

## Data Availability Statement

All data generated or analyzed during this study are included in this article and its online supplementary material files. Additional information, including copies of the questionnaires, are available at https://osf.io/7ep6h/. Further inquiries can be directed to the corresponding author.

## References

1 Kalix P. The pharmacology of khat. Gen Pharmacol. 1984;15(3):179–87.

2 Kalix P. Khat: a plant with amphetamine effects. J Subst Abuse Treat. 1988;5(3):163–9.

3 World Health Organization. Assessment of khat (Catha edulis Forsk). Geneva: WHO; 2006.

4 Toennes SW, Harder S, Schramm M, Niess C, Kauert GF. Pharmacokinetics of cathinone, cathine and norephedrine after the chewing of khat leaves. Br J Clin Pharmacol. 2003; 56(1):125–30.

5 Numan N. Exploration of adverse psychological symptoms in Yemeni khat users by the symptoms checklist-90 (SCL-90). Addiction. 2004;99(1):61–5.

6 Elmi AS. The chewing of khat in Somalia. J Ethnopharmacol. 1983;8(2):163–76.

7 Haile D, Lakew Y. Khat chewing practice and associated factors among adults in Ethiopia: further analysis using the 2011 demographic and health survey. PLoS One. 2015;10(6): e0130460.

8 Alemu WG, Zeleke TA, Takele WW, Mekonnen SS. Prevalence and risk factors for khat use among youth students in Ethiopia: systematic review and meta-analysis, 2018. Ann Gen Psychiatry. 2020;19(1):1–10.

9 Manghi RA, Broers B, Khan R, Benguettat D, Khazaal Y, Zullino DF. Khat use: lifestyle or addiction? J Psychoactive Drugs. 2009;41(1): 1–10.

10 Gebissa E. Scourge of life or an economic lifeline? Public discourses on khat (Catha edulis) in Ethiopia. Subst Use Misuse. 2008;43(6): 784–802.

11 Odenwald M, Neuner F, Schauer M, Elbert T, Catani C, Lingenfelder B, et al. Khat use as risk factor for psychotic disorders: a cross-sectional and case-control study in Somalia. BMC Med. 2005;3(1):1–10.

12 Cox G, Rampes H. Adverse effects of khat: a review. Adv Psychiatr Treat. 2003;9(6):456–63.

13 Mekuriaw B, Zegeye A, Molla A, Hussen R, Yimer S, Belayneh Z. Prevalence of common mental disorder and its association with khat chewing among ethiopian college students: a systematic review and meta-analysis. Psychiatry J. 2020;2020:1462141.

14 Advisory Council on the Misuse of Drugs. Khat: a review of its potential harms to the individual and communities in the UK. London: ACMD; 2013.

15 Thomas S, Williams T. Khat (Catha edulis): a systematic review of evidence and literature pertaining to its harms to UK users and society. Drug Sci Policy Law. 2013;1: 2050324513498332.

16 Odenwald M, Klein A, Warfa N. Introduction to the special issue: the changing use and misuse of khat (Catha edulis) – tradition, trade and tragedy. J Ethnopharmacol. 2010;132(3): 537–9.

17 Mihretu A, Fekadu A, Habtamu K, Nhunzvi C, Norton S, Teferra S. Exploring the concept of problematic khat use in the Gurage community, South Central Ethiopia: a qualitative study. BMJ Open. 2020;10(10):e037907.

18 Mihretu A, Nhunzvi C, Fekadu A, Norton S, Teferra S. Definition and validity of the construct "Problematic Khat Use": a systematic review. Eur Addict Res. 2019;25(4):161–72.

19 Annaheim B. Who is smoking pot for fun and who is not? An overview of instruments to screen for cannabis-related problems in general population surveys. Addict Res Theor. 2013;21(5):410–28.

20 Fiellin DA, Reid MC, O'Connor PG. Screening for alcohol problems in primary care: a systematic review. Arch Intern Med. 2000; 160(13):1977–89.

21 Duresso SW, Matthews AJ, Ferguson SG, Bruno R. Is khat use disorder a valid diagnostic entity? Addiction. 2016;111(9):1666–76.

22 Duresso SW, Matthews AJ, Ferguson SG, Bruno R. Using the severity of dependence scale to screen for DSM-5 khat use disorder. Hum Psychopharmacol. 2018;33(2)–e2653.

23 Mihretu A, Fekadu A, Norton S, Habtamu K, Teferra S. Development of a problematic khat use screening tool in the Gurage community, South-central Ethiopia: classical test theory and item response theory analysis. 2021.

24 Kish L. A procedure for objective respondent selection within the household. J Am Stat Assoc. 1949;44(247):380–7.

25 Kline RB. Principles and practice of structural equation modeling. Guilford Publications; 2015.

26 Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. Psychology Press; 2004.

27 Finch WH, Bronk KC. Conducting confirmatory latent class analysis using mplus. Struct Equ Model. 2011;18(1):132–51.

28 Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med. 1996;3(9):895–900.

29 Association Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub; 2013.

30 Duresso SW, Matthews AJ, Ferguson SG, Bruno R. Is khat use disorder a valid diagnostic entity? Addiction. 2016;111(9):1666–76.

31 Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry. 2010; 32(4):345–59.

32 Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, et al. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. J Affect Disord. 2015;186:32–9.

33 Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, et al. Validity of the patient health questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Res. 2013;210(2):653–61.

34 Üstün TB, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, et al. Developing the world health organization disability assessment schedule 2.0. Bull World Health Organ. 2010;88:815–23.

35 Habtamu K, Alem A, Medhin G, Fekadu A, Dewey M, Prince M, et al. Validation of the World Health Organization Disability Assessment Schedule in people with severe mental disorders in rural Ethiopia. Health Qual Life Outcomes. 2017;15(1):64.

36 Gold LH. DSM-5 and the assessment of functioning: the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). J Am Acad Psychiatry Law. 2014;42(2): 173–81.

37 Bøen H, Dalgard OS, Bjertness E. The importance of social support in the associations between psychological distress and somatic health problems and socio-economic factors among older adults living at home: a cross sectional study. BMC Geriatr. 2012;12(1):27.

38 Fekadu A, Medhin G, Selamu M, Hailemariam M, Alem A, Giorgis TW, et al. Population level mental distress in rural Ethiopia. BMC Psychiatry. 2014;14(1):194.

39 Motrico E, Moreno-Küstner B, de Dios Luna J, Torres-González F, King M, Nazareth I, et al. Psychometric properties of the list of threatening experiences: LTE and its association with psychosocial factors and mental disorders according to different scoring methods. J Affect Disord. 2013;150(3):931–40.

40 Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. Addiction. 1993;88(6):791–804.

41 Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the alcohol use disorders identification test: guidelines for use in primary health care. World Health Organization; 1992.

42 Zewdu S, Hanlon C, Fekadu A, Medhin G, Teferra S. Treatment gap, help-seeking, stigma and magnitude of alcohol use disorder in rural Ethiopia. Subst Abuse Treat Prev Policy. 2019;14(1):4.

43 Gebreyesus SH, Lunde T, Mariam DH, Woldehanna T, Lindtjørn B. Is the adapted Household Food Insecurity Access Scale (HFIAS) developed internationally to measure food insecurity valid in urban and rural households of Ethiopia? BMC Nutr. 2015; 1(1):1–10.

44 Maes KC, Hadley C, Tesfaye F, Shifferaw S. Food insecurity and mental health: surprising trends among community health volunteers in Addis Ababa, Ethiopia during the 2008 food crisis. Soc Sci Med. 2010;70(9):1450–7.

45 Hearon BA, Pierce CL, Björgvinsson T, Fitzmaurice GM, Greenfield SF, Weiss RD, et al. Improving the efficiency of drug use disorder screening in psychiatric settings: validation of a single-item screen. Am J Drug Alcohol Abuse. 2015;41(2):173–6.

46 Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3):297–334.

47 Milfont TL, Fischer R. Testing measurement invariance across groups: applications in cross-cultural research. Int J Psychol Res. 2010;3(1):111–30.

48 Christensen KB, Makransky G, Horton M. Critical values for Yen's Q 3: identification of local dependence in the Rasch model using residual correlations. Appl Psychol Meas. 2017; 41(3):178–94.

49 Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient. 2014;7(1):23–35.

50 Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. J Educ Meas. 1990;27(4): 361–70.

51 Ferguson SL, Moore EWG, Hull DM. Finding latent groups in observed data: a primer on latent profile analysis in Mplus for applied researchers. Int J Behav Dev. 2020;44(5):458–68.

52 Magidson J, Vermunt JK. A nontechnical introduction to latent class models. Statistical Innovations White Paper; 2002. Vol. 1; p. 15.

53 Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Modeling. 2007;14(4):535–69.

54 Guo J, Collins LM, Hill KG, Hawkins JD. Developmental pathways to alcohol abuse and dependence in young adulthood. J Stud Alcohol. 2000;61(6):799–808.

55 Collins LM, Graham JW, Rousculp SS, Fidler PL, Pan J, Hansen WB. Latent transition analysis and how it can address prevention research questions. NIDA Res Monogr. 1994; 142:81.

56 Vermunt JK, Magidson J. Latent class cluster analysis. Appl Latent Class Anal. 2002;11(89–106):89.

57 Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika. 2001;88(3):767–78.

58 Wang Z, Jegelka S. Max-value entropy search for efficient Bayesian optimization. arXiv. 2017. Submitted.

59 Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif. 1996;13(2):195–212.

60 Muthén B, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000;24(6): 882–91.

61 Weller BE, Bowen NK, Faubert SJ. Latent class analysis: a guide to best practice. J Black Psychol. 2020;0095798420930932.

62 Vermunt JK, Magidson J. Upgrade manual for Latent GOLD 5.1. Belmont (MA): Statistical Innovations; 2016.

63 Stata Corporation. College Station (TX).

64 Arbuckle J. IBM SPSS Amos 23 user's guide. Amos Development Corporation; 2019. Vol. 2014.

65 Lauritsen J, Bruus M. EpiData-An extended tool for validated data-entry and documentation of data. Odense, Denmark: The EpiData Association; 2003.

66 Gryczynski J, McNeely J, Wu LT, Subramaniam GA, Svikis DS, Cathers LA, et al. Validation of the TAPS-1: a four-item screening tool to identify unhealthy substance use in primary care. J Gen Intern Med. 2017;32(9):990–6.

67 WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. Addiction. 2002;97(9):1183–94.

68 Saha DT, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the national epidemiologic survey on alcohol and related conditions. Psychol Med. 2006;36(7):931.

69 Jackson K, Bucholz KK, Wood PK, Steinley D, Grant JD, Sher KJ. Towards the characterization and validation of alcohol use disorder subtypes: integrating consumption and symptom data. Psychol Med. 2014;44(1):143–59.

70 Cleveland MJ, Mallett KA, White HR, Turrisi R, Favero S. Patterns of alcohol use and related consequences in non-college-attending emerging adults. J Stud Alcohol Drugs. 2013; 74(1):84–93.

71 Craft S, Winstock A, Ferris J, Mackie C, Lynskey MT, Freeman TP. Characterising heterogeneity in the use of different cannabis products: latent class analysis with 55,000 people who use cannabis and associations with severity of cannabis dependence. Psychol Med. 2020;50(14):2364–73.

72 Myrseth H, Notelaers G. A latent class approach for classifying the problem and disordered gamers in a group of adolescence. Front Psychol. 2018;9:2273.

73 Tomczyk S, Isensee B, Hanewinkel R. Latent classes of polysubstance use among adolescents-a systematic review. Drug Alcohol Depend. 2016;160:12–29.

74 Pontes HM, Király O, Demetrovics Z, Griffiths MD. The conceptualisation and measurement of DSM-5 internet gaming disorder: the development of the IGD-20 test. PLoS One. 2014;9(10):e110137.

75 Kassel JD, Stroud LR, Paronis CA. Smoking, stress, and negative affect: correlation, causation, and context across stages of smoking. Psychol Bull. 2003;129(2):270.

76 Dierker LC, Vesel F, Sledjeski EM, Costello D, Perrine N. Testing the dual pathway hypothesis to substance use in adolescence and young adulthood. Drug Alcohol Depend. 2007;87(1):83–93.

77 Sofis MJ, Borodovsky JT, Pike CK, Liu L, Jacobson NC, Budney AJ. Sifting through the weeds: relationships between cannabis use frequency measures and delay discounting. Addict Behav. 2020;112:106573.

78 Alem A, Kebede D, Kullgren G. The prevalence and socio-demographic correlates of khat chewing in Butajira, Ethiopia. Acta Psychiatr Scand Suppl. 1999;397:84–91.

79 Mihretu A, Teferra S, Fekadu A. What constitutes problematic khat use? An exploratory mixed methods study in Ethiopia. Subst Abuse Treat Prev Policy. 2017;12(1):17–2.

80 Annaheim B, Legleye S. Short instruments to screen for "problematic" cannabis use in general population surveys, in handbook of cannabis and related pathologies. Elsevier; 2017. p. e168–84.