The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility

WHO ASSIST Working Group
Correspondence to: Thomas F. Babor PhD University of Connecticut Health Center Department of Community Medicine and Health Care 263 Farmington Avenue Farmington, CT 06030 6325 USA

ABSTRACT

Aims

The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) was developed for the World Health Organization (WHO) by an international group of substance abuse researchers to detect psychoactive substance use and related problems in primary care patients. This report describes the new instrument as well as a study of its reliability and feasibility.

Setting

The study was conducted at participating sites in Australia, Brazil, Ireland, India, Israel, the Palestinian Territories, Puerto Rico, the United Kingdom and Zimbabwe. Sixty per cent of the sample was recruited from alcohol and drug abuse treatment facilities; the remainder was drawn from general medical settings and psychiatric facilities.

Methods

The study was concerned primarily with test item reliability, using a simple test–retest procedure to determine whether subjects would respond consistently to the same items when presented in an interview format on two different occasions. Qualitative and quantitative data were also collected to evaluate the feasibility of the screening items and rating format.

Participants

A total of 236 volunteer participants completed test and retest interviews at nine collaborating sites. Slightly over half of the sample (53.6%) was male. The mean age of the sample was 34 years and they had completed, on average, 10 years of education.

Results

The average test–retest reliability coefficients (kappas) ranged from a high of 0.90 (consistency of reporting 'ever' use of substance) to a low of 0.58 (regretted what was done under influence of substance). The average kappas for substance classes ranged from 0.61 for sedatives to 0.78 for opioids. In general, the reliabilities were in the range of good to excellent, with the following items demonstrating the highest kappas across all drug classes: use in the last 3 months, preoccupied with drug use, concern expressed by others, troubled by problems related to drug use, intravenous drug use. Qualitative data collected at the end of the retest interview suggested that the questions were not difficult to answer and were consistent with patients' expectations for a health interview. The data were used to guide the selection of a smaller set of items that can serve as the basis for more extensive validation research.

Conclusion

The ASSIST items are reliable and feasible to use as part of an international screening test. Further evaluation of the screening test should be conducted.

INTRODUCTION

According to the United Nations Drug Control Programme (UNDCP 1997), illicit drug consumption has increased throughout the world in recent years, as estimated from emergency room visits, drug-related mortality reports and arrests of drug users. Based on unofficial UNDCP estimates, the annual global prevalence of illicit drug consumption is in the range of 3–4% of the world population, which is far below the estimated consumption of the two primarily legal substances, alcohol (50%) and tobacco (20%), but nevertheless of sufficient magnitude to warrant a public health approach to early intervention. A key ingredient of a public health approach is the availability of a reliable and valid screening test.

The development of any screening test on an international level must take into account the diverse nature of psychoactive substances and of the users themselves. Much has been learned from recent research about screening for substance abuse, and these findings have provided a sound empirical basis for the design of a new screening test (McPherson & Hersch 2000; Babor 2002). Two different approaches have been developed to screen for drug abuse: self-report procedures and biochemical tests. The first, exemplified by the Drug Abuse Screening Test (DAST, Skinner 1982), consists of direct questions about substance use and related problems. Another self-report screening approach, illustrated by the MacAndrew Alcoholism Scale (MacAndrew 1965), consists of indirect questions that measure personality traits associated with substance misuse. A variation on the use of direct screening questions focusing on specific substances is the use of conjoint screening items, which inquire simultaneously about alcohol and drugs. One such instrument is the CAGE alcoholism screening test adapted to include drugs (CAGE-AID, Brown & Rounds 1995).

A second type of screening method for drug abuse is biological tests. Drug screening through such methods as urinalysis, hair testing and saliva tests, when performed by someone with proper training, is often the quickest, most accurate way to detect recent drug use (Wolff *et al.* 1999).

The available self-report screening tests have a number of limitations from an international perspective (Babor 2002). First, most tests have been developed and validated in the United States, with little evidence that they are sensitive or specific in other cultures. Secondly, many of these tests have not been subject to extensive validation research, and several have not been evaluated favorably. Thirdly, most of the adult screening tests have been designed for case finding, not to identify risk factors. These tests typically avoid direct questions about specific drugs, focusing instead on the experience of problems associated with any drug use in the past. Such an approach may be susceptible to imprecision and response bias. Fourthly, subtle or disguised screening tests do not appear to be sufficiently sensitive to identify active cases, although they may be useful in screening for risk factors. Finally, biological screening methods also have limitations (Wolff *et al.* 1999), such as the problems of handling body fluids and the need for scrupulous hygienic procedures. Cost, invasiveness, lack of sensitivity and other problems limit the usefulness of biological measures as a sole approach to screening, especially in developing countries.

Based on similar considerations, a critical review of current drug abuse screening techniques (US Preventive Services Task Force 1998) concluded that there is insufficient evidence to recommend for or against routine screening using currently available, standardized self-report questionnaires or biologial assays. Given the prevalence and consequences of drug use throughout the world, it appears that there is a need for an international screening test for drug abuse that is reliable, valid, flexible, comprehensive and inexpensive. Further work is needed therefore in the development and validation of accurate screening instruments, especially instruments suitable for use in primary care in areas of the world where drug abuse is prevalent but difficult to detect. To address this need, the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) was developed recently for the World Health Organization (WHO) by an international group of substance abuse researchers for field testing in representative member states in different parts of the world. This report describes the results of an international study to test the feasibility and reliability of the new screening test.

Please use the following citation: WHO ASSIST Working Group (2002), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility, (Author postprint) in *Addiction,* 97(9), 1183-1194, [Accessed: (date) from www.drugsandalcohol.ie

METHODS

Working in close collaboration with the WHO Department of Mental Health and Substance Dependence, the study was conducted at participating sites in Australia, Brazil, India, Ireland, Israel, the United Kingdom, Zimbabwe, the Palestinian Territories and Puerto Rico, a territory of the United States. The centers were chosen for their ability to provide access to individuals with different substance use patterns. The study was conducted in two stages. The first stage was devoted to planning and instrument development. The second stage involved a test–retest exercise and the evaluation of the test's feasibility.

Planning and development of ASSIST 1.0

Ideally, an international screening test should be adaptable to different cultures, languages and settings. It should be flexible enough to identify different substances and different patterns of substance use. It should also be capable of serving different purposes, such as case finding as well as screening for risk factors. To meet these standards, an international group of substance abuse researchers was convened to design a preliminary screening procedure that could be subject to testing for reliability and feasibility. Based on a critical review of the literature in screening for drug abuse (Babor 2002), the group developed ASSIST 1.0 according to the following guidelines: (1) simple instructions that explain the purpose and the nature of the screening questions; (2) the inclusion of questions about alcohol and tobacco in order to present the drug items in the context of a more general health and lifestyle screening interview; (3) the use of specific drug classes that allow differentiation across the full range of abuse liability and dependence potential associated with different types of psychoactive substances; and (4) direct questions about the current and past use of each substance, as well as problems and dependence symptoms associated with substance use.

After compiling a comprehensive inventory of test items and response formats that had been used in previous interviews and questionnaires, candidate items were constructed for each of the proposed content domains. The 12 items selected by consensus for initial evaluation provided ample coverage of the content domains considered most relevant to screening: life-time and recent substance use, dependence symptoms, substance-related problems and intravenous use. It was decided to include alcohol and tobacco products among the substance classes in order to integrate drug screening with substances that have achieved greater acceptance in primary care as targets of screening programs. Substances other than tobacco and alcohol were differentiated into eight drug classes (rather than combining them into a single drug category). This was considered important because of the wide variability among these substances in abuse liability and dependence potential.

ASSIST 1.0 began with an initial screening item that asked about life-time use of commonly used substances within the following 10 categories: tobacco products, alcoholic beverages, cannabis, cocaine, stimulants, inhalants, sedatives/hypnotics, hallucinogens, opioids and 'other drugs.' If the respondent reported no psychoactive substance use, the interview was terminated. If the respondent admitted to life-time use of one or more substances, the remaining questions were asked only for those substances endorsed in the initial screening question. The 11 additional questions were selected from a large pool of items that had been used in previously developed scales to measure the frequency of substance use, dependence symptoms, substance-related problems, and injection drug use. It was expected that a smaller number of items would be selected for inclusion in the revised ASSIST interview upon completion of the reliability study.

Following the question about life-time use, the second question asked about drug use during the 'past 3 months'. The drug classes in this question were rated on a five-point frequency scale ranging from 'never' (in the past three months) to 'daily'. This question provided critical information about the substances most relevant to the respondent's current health status.

If none of the substances had been used in the past 3 months, the interviewer skipped to questions 9–12 about problems and use patterns that occurred prior to the past 3 months. These questions inquired about a past history of harmful use or dependence that may increase the risk of future problems even in the absence of current use.

If any substance was used once or more during the past 3 months, the remaining questions were asked. Question 3 asked about preoccupation with the substance or substances used in the past 3 months. This was intended as a measure of the salience of substance use, an indicator of psychological dependence. Question 4 asked about worry or concern about substance use, a possible sign of impaired control. Question 5 queried about problems with family, work or school. This is a measure of harmful use. Question 6 dealt with inability to manage routine role obligations, such as care of children or employment responsibilities. This was another measure of salience of drug use. Question 7 examined guilt associated with drug use. Question 8 asked about use that exceeded intended limits, another measure of impaired control. The remaining questions inquired about life-time and current problems (questions 9 and 11), as well as prior attempts at control (question 10) and intravenous drug use (question 12).

Some of the drugs listed may have been prescribed by a doctor (such as amphetamines, sedatives or pain medications). If a participant reported the use of a prescription medication, interviewers determined if the drug was taken for reasons other than a doctor prescribed, or was taken more frequently or at higher doses than prescribed. Interviewers were instructed to code only those drugs that were 'abused', not those taken as prescribed for medical reasons.

Reliability and feasibility studies

The second stage of the project involved the following sequence of activities: (1) the training of at least two interviewers at each site to administer ASSIST 1.0; (2) the recruitment of approximately 25 individuals per center who met inclusion and exclusion criteria; (3) the evaluation of instrument feasibility and test–retest reliability; and (4) further revision of the new drug screening instrument based on the study findings.

The aims of the reliability study were: (1) to determine specific reliability of the preliminary items in order to make revisions in the content of the items, the length of the screening test, and the recommended response categories; and (2) to identify possible sources of response error in the instructions and format that required modification before a more systematic validity study could be conducted.

The aim of the feasibility study was to evaluate the applicability and acceptability of the new screening procedure with different types of patients, in different settings, in different cultures, across a variety of substances. Patients and interviewers were asked to rate the proposed items in terms of their comprehensibility, cultural appropriateness and ability to elicit honest answers. Focus group exercises were also conducted at several sites.

Research design

Test–retest reliability is the comparison of ratings made during an initial interview, with independent ratings made in a second interview. The second interview (retest) was conducted by an interviewer blind to the results of the first interview. The period of time between the test and retest assessments was between one day and three days. To minimize practice effects and respondents' desire to provide identical information, the interviewer explained that during the second interview they should answer the questions directly and not attempt to recall their answers from the previous occasion.

After the second (retest) interview was completed, the retest interviewer (or a third independent interviewer) asked the respondent to answer questions from a separate discrepancy interview that was designed to evaluate the feasibility of the ASSIST questions and procedures. The discrepancy interview began by reviewing with the respondent both the initial and second ratings on the designated items. If discrepancies occurred, the interviewer explored the reasons that gave rise to the discrepancy (in the respondent's view), and noted the respondent's responses to these questions. Following the discrepancy interview, the interviewer provided both qualitative and quantitative ratings of cultural appropriateness, comprehensibility and defensiveness. The latter ratings were designed to evaluate the potential bias associated with culturally sensitive questions.

Collaborating sites were asked to conduct at least one focus group to explore several issues surrounding the use of screening tests in primary care in that particular country. Focus group participants included interviewers, investigators and research assistants. They were asked to

describe the reactions of different types of patients to the screening procedure, to comment on similarities and differences observed between samples, to provide their own subjective reactions to the procedures, to identify problems with administration of items and to evaluate the accuracy of the screening method.

Translation and cultural adaptations

Prior to subject recruitment, investigators at each site compiled a list of culture-specific names of common drugs of abuse in each country. The most commonly abused substances were listed for the country, as well as drugs most often associated with treatment and those most associated with legal problems. This information was used for the local adaptation of the drug classes (including colloquial names and examples) of the ASSIST screening interview used in the test–retest exercise. The study questionnaires, participant instructions, and response scales were translated into the local languages by at least two bilingual translators. After the materials were translated from English to the local language, any differences were discussed by the two translators. Corrections were made before the final translation was completed.

Recruitment procedures

Each collaborating center recruited approximately 25 volunteers. Forty per cent of the sample was recruited from general medical, primary care or community settings while the remaining participants were drawn from specialized alcohol and drug treatment services, primarily residential settings where patients could be reinterviewed conveniently several days later. A total of 236 sets of test–retest interviews were completed at the nine collaborating sites. Sites each contributed between 14 and 30 sets of test–retest exercises.

The rationale for these recruitment procedures was both practical and methodological. On the practical side, cross-national instrument development research is expensive and resource-intensive, which limited the number of subjects to the minimum required for adequate statistical power at the level of the total sample rather than at the individual site level. Drug treatment sites were chosen because they provide access to a population of drug users, and demonstrate ample variability across multiple substances in the frequency of substance use and related problems. The remaining sites were chosen to provide access to facilities where a range of substance users, particularly those with low and moderate levels of health risk, could be recruited to add variability to the measures under consideration.

The following exclusion criteria were used to screen out inappropriate study participants: (1) communication difficulties (e.g. language problems, deafness); (2) cognitive impairment or mental retardation; (3) severe behavior disturbance, psychotic symptoms, uncooperativeness; (4) drug and alcohol intoxication or withdrawal; (5) likelihood that respondent would not be available for the second interview; and (6) age of respondent was under 18 or above 60.

Interviewers

Interviewers *(*2–4 per site) were selected on the basis of their familiarity with drugs of abuse. Seventy-one per cent were female. Their average age was 32 years. They had approximately 15 years of education and many had advanced degrees (14% with Masters, 29% with PhDs, and 5% with MDs). One-fourth of the interviewers were working as researchers, one-third were employed primarily in the alcohol and drug treatment field and the remainder worked in medical, psychiatric or other facilities.

Characteristics of the sample

Slightly over half of the sample (53.6%) was male and 33.7% were married or cohabiting. Sixty-one per cent of the participants were categorized as 'unemployed', including students (1.9%), homemakers (10.5%) and those with disabilities (4.8%). Twenty-five per cent of the sample were employed full time and another 14% were employed part-time. The mean age of the respondents was 34 years. They had completed, on average, 10 years of education. Over half of the sample (59.7%) was recruited from alcohol and drug abuse treatment facilities while the remainder was drawn from general medical settings and psychiatric facilities.

Data analysis

The data were examined according to question stem (i.e. the main subject of the question, regardless of the type of substance), substance class and recruitment setting in order to provide recommendations for improving the ASSIST instrument. Qualitative data (the debriefing interview and results of the focus groups) were summarized by each center in English. Because of the small sample size and the study's focus on cross-national representativeness, site-specific analyses were not conducted. The analysis strategy consisted of: (1) computing kappa coefficients for each item in each drug class; (2) comparing average kappas across drugs to identify the most reliable items; (3) computing correlations between frequency of substance use (within each drug class) and the frequency of symptoms; (4) comparing internal consistency reliability coefficients (Cronbach's alphas) for the entire set of 11 items (excluding the life-time use question and the alcohol and tobacco drug classes); (5) examining the reduction in alpha if an item is deleted; (6) reviewing the qualitative data collected through the debriefing interview and focus groups with interviewers; (7) choosing eight items (referred to as ASSIST 2.0) on the basis of high kappas, high correlations with drug use frequency, strong item-total scale correlations, face validity, lack of overlap with other items and support from qualitative data.

RESULTS

Reliability

On average, the time for administering the ASSIST 1.0 was 16 minutes, while the average retest time was 17.5 minutes. As expected, subjects recruited from alcohol and drug treatment facilities took significantly longer to complete the interview than subjects recruited from general medical settings (test time: 17.88 versus 13.26, $t = 4.87$, $P < 0.001$; retest time: 19.70 versus 14.33, $t = 4.06$, $P < 0.001$). The average time between the test and retest interview was 2.16 days and ranged from 1.32 in Sao Paulo to 3.28 in Harare.

Test–retest coefficients of agreement (kappa-values) for both question stem and drug category are shown in Table 1. Prevalence rates for each drug category are noted in the first column of the table. Four categories (cannabis, cocaine, sedatives and opioids) had current prevalence rates ranging from 23% (current use of cocaine) to 42% (current use of cannabis). These rates are sufficient to place confidence in the kappas. Drug classes with lower prevalence rates may have biased kappas, or coefficients may not have been computed at all. The table provides a summary of the test–retest analyses by averaging kappas across items and across substances.

Table 1 Test–retest kappa statistics by question and substance class (n=236).

Substance class (prevalence rate, ever/current)	ASSIST 1.0 Questions: Ever used Q1	Used last 3 months Q2	Preoccupied Q3	Worried Q4	Problems Q5	Neglect Q6	Regrets Q7	Used more Q8	Concerns Q9	Cut down Q10	Troubled by problems Q11	i.v. use Q12	Average kappa
(a) Tobacco products (93.2/85.6)	**0.97**	**0.95**	**0.67**	**0.64**	**0.50**	**0.24**	**0.51**	**0.53**	**0.77**	**0.68**	**0.58**	**(b)**	**0.64**
(b) Alcoholic beverages (95.8/72.9)	0.75	0.85	0.74	0.71	0.60	0.68	0.76	0.74	0.75	0.68	0.70	(b)	0.72
(c) Cannabis (76.7/42.4)	0.95	0.76	0.72	0.66	0.52	0.64	0.43	0.58	0.63	0.54	0.58	(b)	0.64
(d) Cocaine or crack (41.9/22.9)	0.96	0.75	0.72	0.73	0.73	0.67	0.64	0.72	0.79	0.74	0.77	0.94	0.76
(e) Stimulants (44.1/15.7)	0.91	0.80	0.76	0.67	0.63	(a)	0.60	0.77	0.73	0.67	0.68	0.97	0.74
(f) Inhalants (18.6/2.5)	0.91	0.72	0.80	(a)	(a)	(a)	(a)	0.86	0.56	0.65	0.76	(b)	(a)
(g) Sedatives (56.8/32.6)	0.75	0.66	0.61	0.56	0.48	0.59	0.58	0.59	0.57	0.59	0.65	0.68	0.61
(h) Hallucinogens (36.0/5.5)	0.96	(a)	0.53	0.33	(a)	(a)	(a)	(a)	0.63	0.72	0.65	(b)	(a)
(i) Opioids (55.9/39.4)	0.94	0.83	0.77	0.82	0.80	0.76	0.61	0.64	0.81	0.78	0.79	0.84	0.78
(j) Other drugs (21.2/14.4)	0.81	0.82	0.91	0.81	0.70	0.71	0.60	0.59	0.80	0.73	0.72	0.70	0.74
Average kappa value for drug classes c–j	0.90	0.76	0.73	0.65	0.64	0.67	0.58	0.68	0.69	0.68	0.70	0.83	

a) = insufficient data to compute Kappa coefficients. (b) = Question does not apply to that substance

The average kappas ranged from a high of 0.90 (consistency of reporting 'ever' use of substance) to a low of 0.58 (regretted what was done under influence of substance). In general, the reliabilities were in the range of good to excellent, with the following items demonstrating the highest kappas across all drug classes: Q.2 (use in the last 3 months, κ = 0.76), Q.3 (preoccupied with drug use, κ = 0.73), Q.9 (concern expressed by others, κ = 0.69), Q.11 (troubled by problems related to drug use, κ = 0.70) and Q.12 (i.v. drug use, κ = 0.83).

Using the discrepancy interview, information was obtained about the reasons for differences between test and retest answers, when they occurred. The most commonly mentioned explanations were respondent errors such as 'I didn't understand the question' (24%), 'I was not paying attention on one occasion' (18%), 'I couldn't remember the answer at the time' (15%), and 'I remembered this experience between interviews' (14%). Twenty-three per cent of the errors were attributed to interviewer error: 'My answers were not different. They seem to have been misunderstood by one of the interviewers' (13%) and 'one interviewer probed more than the other' (5%).

The next set of analyses evaluated internal consistency reliability (alpha coefficients) and item correlations with current frequency of substance use. The results were used to identify items that contributed to the overall coherence of scales within each drug class. These analyses were conducted using the four drug categories with the highest prevalence rates and the most acceptable distributional properties (cannabis, cocaine, sedatives, opioids). The results are summarized in Table 2, along with the data from alcohol and tobacco. The

Please use the following citation: WHO ASSIST Working Group (2002), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility, (Author postprint) in *Addiction,* 97(9), 1183-1194, [Accessed: (date) from www.drugsandalcohol.ie

correlations are high for most items across all substances (except for tobacco). Because of the overlap in item content, it is not unexpected that the alpha coefficients for these scales are also high. The items, particularly those measuring social consequences, performed least well with respect to tobacco use.

Table 2 Alpha statistics and correlations (*r* s) between symptom frequency and frequency of substance use within six substance classes**.**

	Substance class					
Symptom	**Cannabis (*n* = 100)**	**Cocaine (*n* = 54)**	**Sedatives (*n* = 77)**	**Opioids (*n* = 93)**	**Tobacco (*n* = 202)**	**Alcohol (*n* = 172)**
Preoccupied[a] *r*	**0.74**	**0.51**	**0.58**	**0.70**	**0.37**	**0.59**
Worried[a] *r*	0.31	0.55	0.57	0.52	0.16	0.46
Problems[a] *r*	0.37	0.52	0.38	0.65	0.07	0.42
Neglect[a] *r*	0.37	0.59	0.39	0.49	0.04	0.45
Regret[a] *r*	0.13	0.55	0.29	0.41	0.05	0.44
More than intended[a] *r*	0.38	0.57	0.37	0.49	0.16	0.57
Concern[a] *r*	0.29	0.30	0.38	0.50	0.02	0.39
Cut down[b] *r*	0.15	0.32	0.36	0.22	-0.03	0.41
Troubled[b] *r*	0.30	0.54	0.30	0.31	0.08	0.38
i.v. drug use[b] *r*	(c)	0.28	0.20	-0.02	(c)	(c)
Alpha[d]	0.85	0.91	0.87	0.85	0.73	0.92

[a] Item rated on a five-point scale (0 = never, 1 = less than once a month, 2 = monthly, 3 = weekly, 4 = daily).
[b] Item rated on a three-point scale (0 = no, never, 1= yes, but not in the past 3 months, 2 = yes, in the past 3 months).
[c] Not applicable.
[d] Substance-specific internal consistency reliabilities (Cronbach's alphas) for the 10 candidate items.

Feasibility data

The debriefing interview, conducted at the end of the test–retest exercise, indicated that 96% of the participants enjoyed being interviewed, and 81% thought that the length of the interview was 'just right'. A large majority (78%) of the respondents indicated that none of the questions were difficult to understand, and virtually all participants (98%) thought the questions were not offensive. In general, respondents had little difficulty understanding the questions and considered them to be appropriate.

The debriefing interview also included two items recommended by Sudman & Bradburn (1982) to measure potential response bias in the context of a health survey. In general, the greater the perceived threat associated with a question, the greater the likelihood of denial or minimization. Perceived threat can be detected by projective questions asking respondents to estimate the discomfort and dishonesty 'other people' are likely to experience answering the same questions to which they have just responded. Respondents who report that a question would make 'most people' very uneasy are more likely to under-report than other respondents are. Respondents who believe that other people are not likely to answer a question honestly are more likely to be dishonest themselves. Respondents were therefore asked to rate on a five-point scale how comfortable other people would be answering the ASSIST questions and how honest they thought other people would be.

Table 3 shows that participants believed that 'most people' would be more comfortable answering questions about tobacco and alcohol (4.00 and 3.27, respectively, on a scale of 1–5: 1 = very uncomfortable, 5 = very comfortable) than questions about cocaine and opioids (2.29 and 2.44, respectively).

Please use the following citation: WHO ASSIST Working Group (2002), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility, (Author postprint) in *Addiction,* 97(9), 1183-1194, [Accessed: (date) from www.drugsandalcohol.ie

Table 3 Respondents' (*n* =134) ratings of ASSIST questions in terms of discomfort likely to be experienced by others, [a] susceptibility to dishonest responding [b] and perceived importance of the screening information to their health care provider. [c]

Substance	Comfort[a] X	(SD)	Honesty[b] X	(SD)	Importance[c] X	(SD)
Tobacco	4.00	(1.25)	3.74	(1.30)	4.64	(0.94)
Alcohol	3.27	(1.41)	3.30	(1.30)	4.62	(0.96)
Marijuana	3.03	(1.42)	2.81	(1.30)	4.51	(1.03)
Cocaine	2.29	(1.43)	2.41	(1.26)	4.66	(0.90)
Opioids	2.44	(1.42)	2.46	(1.27)	4.67	(0.87)
Other substances	2.79	(1.34)	2.69	(1.23)	4.55	(0.96)

[a] How comfortable are others answering this survey? (1 = very uncomfortable, 5 = very comfortable).
[b] How honest are others when telling their health care provider about health habits? (1 = very honest, 5 = very dishonest).
[c] How important is it that your health care provider know about your health habits? (1 = very uncomfortable, 5 = very comfortable).

Three-way analysis of covariance was used to test the effects of substance class, recruitment site (primary care vs. drug treatment) and gender, using the respondent's age as a covariate. Site effects were not tested because of the small sample sizes. Table 3 shows that respondents rated some substances (e.g. tobacco) in the range of 'comfortable' (M = 4.00), whereas other substances (e.g. cocaine) were rated more in the uncomfortable range (M = 2.29). There were significant differences across substances in the comfort ratings ($F_{5.630}$ = 12.05, P < 0.001), and there was a significant substance by recruitment site interaction effect ($F_{5,630}$ = 2.33, P < 0.05). General medical patients rated 'others' as being more comfort-able answering questions about alcohol and tobacco, whereas drug treatment patients rated others as being more comfortable with questions about marijuana and cocaine. The results were similar in response to a question asking 'how honest are others when telling their health care provider about' these substances. Respondents believed that 'other people' would be most honest answering questions about tobacco and alcohol, and least honest about cocaine and heroin ($F_{5,645}$ = 11.3, P < 0.001). Although there were no gender differences, respondents recruited from drug treatment settings had significantly higher honesty ratings than those recruited from primary care ($F_{1,129}$ = 4.23, P < 0.05).

Finally, participants were asked to rate how important it was for their health care provider to know about the substances listed in the ASSIST. Respondents rated all of the substances on average as very important for their provider to know about, with no significant differences by substance class, gender or recruitment site.

Feasibility was also investigated by means of questions answered by each of the interviewers. The results indicated that the interviewers thought that participants were interested in the interview (97%), were not offended by the questions (100%), and were not responding in the negative to shorten the interview (95%). None of the interviewers perceived that participants were withholding information. Other questions asked interviewers to identify unclear or confusing items in the ASSIST. Comments were varied and covered a range of issues such as redundancy, vocabulary problems and time frame issues (e.g. confusion between 'last 3 months' versus 'life-time'). Two items, 'preoccupied with thoughts about using' and 'used more than intended', were identified as confusing. Seventy-seven per

Please use the following citation: WHO ASSIST Working Group (2002), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility, (Author postprint) in *Addiction,* 97(9), 1183-1194, [Accessed: (date) from www.drugsandalcohol.ie

cent of the interviewers indicated that the interview was 'easy' or 'very easy' to administer. Only 3% said that it was very difficult to conduct.

Focus group exercises were conducted and site reports were completed to collect qualitative information on feasibility. Three sites (Adelaide, Sao Paulo and Beer Sheva) submitted focus group reports. Two reports identified problems with the instructions for 'prescription drug use' versus 'abuse of drugs'. It was felt that these instructions should be placed at the beginning of the interview and should be clarified so that respondents do not include medical use of drugs when answering the screening questions. Some questions were felt to be too general or 'unfocused' (e.g. questions 3 and 4 'thoughts or concerns'). A suggestion was also made to clarify the frequency categories for the ASSIST items.

Finally, a number of practical problems and logistical issues were identified during the reliability testing and data processing. A consistent problem was drug categorization. Each site was asked to create a culture-specific 'drug card' to list the primary drugs (e.g. valium, librium) for each major drug category (e.g. sedatives). The drug card should have listed the drugs most likely to be abused in that country, but may not necessarily have included every drug possibility. Drugs not listed as examples were consistently coded in the 'other' category when in some cases they should have been included in one of the primary categories. For example, there was confusion about where to list methadone. Another common problem was that prescription drug use was sometimes included as a drug of abuse even when the subject or interviewer noted that the drug was taken only as prescribed.

Assist 2.0

The reliability and feasibility data were used to revise and shorten the ASSIST. Two meetings of the investigators were held to coordinate the revision. The following criteria were applied in revising the instrument on the basis of the evidence obtained from the reliability and feasibility studies: (1) simplicity (the test should be brief and easy to apply in terms of its structure and format); (2) applicability (the test should be compatible with primary health care settings); (3) coverage of key domains (i.e. substance use, problems, dependence symptoms, i.v. use); (4) appropriateness for use with a range of people and problems (i.e. hazardous use; harmful use, dependence); and (5) compatibility with the empirical data (i.e. the items should have high internal consistency and test–retest reliability).

Based on these guidelines, two types of modification were made. First, the length of the test was reduced to eight items. Based on a review of kappa coefficients, item-total correlations, distributional properties, qualitative feedback, correlation with other items, problems with language translations and coverage of relevant content domains, four items were dropped [items 4 (worried or concerned about own use); 7 (feelings of regret); 8 (used more than intended) and 11 (bothered by problems)]. Secondly, comprehension and coverage of some questions were improved. This was accomplished by: (1) changing the wording of three items; (2) revising the wording of response formats; (3) revising site adaptation procedures and instructions; (4) developing scoring procedures; and (5) retaining the structure and reference periods of the original test (life-time, 3 months). The revised questions are listed in Appendix I. These questions now form the basis for the screening questionnaire (ASSIST 2.0) that can be used in further evaluation research.

The final structure and scoring routines for the ASSIST will be derived from a subsequent international validation study. Given the acceptable face validity and distributional properties of the summary scores derived from ASSIST 1.0, it is probable that ASSIST 2.0 will permit the computation of the following measures: (1) life-time drug use (i.e. a count of the positive answers to question 1); (2) current substance use frequency (for each substance indicated in item 1, an estimate of the current frequency of use); (3) current substance-specific

severity (for each substance indicated in question 2, a sum of responses to questions 2–7 provide an estimate of the degree of dependence and problems associated with that substance; and (4) injection drug use risk score (response to question 8). More complete information on the use and scoring of the ASSIST can be obtained from the WHO address provided at the end of this paper.

DISCUSSION

In psychological testing, reliability typically refers to the consistency of response to test items, and can be affected by the test item content, the person tested, the person undertaking the testing and the conditions under which the test is administered. The present study was concerned primarily with test item reliability, using a simple test–retest procedure to determine whether subjects would respond consistently to the same items when presented in an interview format on two different occasions. The analyses using kappa statistics gave evidence of high test–retest reliability, with some variability across drug classes and items. Internal consistency analyses suggested that the 11 original items were highly intercorrelated, but this was to be expected because of the overlap in content among some of the items. In general, these analyses suggest that reliability was acceptably high across all sites.

Interviews with respondents suggested that the questions were not difficult to answer and were consistent with their expectations for a health interview. More importantly, there did not seem to be any confusion about the flow, format and structure of the interview. Most respondents had no criticisms or comments, and when problems occurred, they were caused by specific items and involved only small numbers of respondents. Nevertheless, there was some indication from the projective questions dealing with comfort and honesty (Table 3) that some respondents may perceive the drug questions to be threatening, and that they may minimize or deny drug use as a result. This finding suggests that further validation work on ASSIST 2.0 should be particularly attentive to the detection and minimization of response bias (Babor & Del Boca 1992).

In general, qualitative data from the debriefing interview did not suggest that there were major problems associated with the administration of the ASSIST interview. Many problems mentioned in the open-ended comments were due to the large set of preliminary question stems that included redundant and overlapping items. This problem was resolved easily by reducing the number of items in the revised screening test.

Comments provided by the interviewers corroborate these conclusions. The ASSIST items and format, despite some shortcomings, seemed to be feasible and appropriate, at least under research conditions. Concerns about length and redundancy were expected because the preliminary instrument was designed to be overly inclusive.

The revised version of ASSIST (2.0) is now being released for further validation research and feasibility testing. ASSIST 2.0 is designed to be brief, flexible, comprehensive and appropriate for use in primary care. Although ASSIST 1.0 required approximately 13 minutes, the revised version should take less than 5 minutes for most general practice patients because of the reduced item content and the addition of skipout instructions. Because ASSIST also screens for tobacco and alcohol use, this seems an acceptable amount of time to devote to risk factors that are not only likely to co-occur in the same individuals, but are also likely to require similar clinical interventions.

Unlike other drug abuse screening tests, the ASSIST is designed to provide life-time, as well as current, estimates of substance-related risk. Screening with life-time questions like the Drug Abuse Screening Test or the CAGE-AID provides little information beyond the fact that the patient has used drugs sometime in the past. ASSIST 2.0 provides drug-specific

information about life-time use and current severity, which should be useful for health education and early intervention. The drug-specific information should also be useful when there is an opportunity to coordinate ASSIST data with biological screening results such as hair or urine analysis.

Nevertheless, further validation research is necessary, especially in light of the limitations of the present study. First, the sample size was too small to permit more refined statistical analyses at the site level, and even with the data aggregated across samples, the numbers within certain drug categories were marginal or too small for analysis. Secondly, it was beyond the purpose of the present study to evaluate the concurrent and predictive validity of either ASSIST 1.0 or 2.0, which is a critical prerequisite to the eventual use of the test in primary care settings. The eight final items were selected on the basis of their association with drug use frequency, and not with ICD-10 dependence or harmful use. Thirdly, the preliminary version of ASSIST used in this investigation was not evaluated under typical conditions of primary care. Sixty per cent of the participants were patients in drug treatment settings, and only a small number of the interviewers were primary care health professionals. It is possible that ASSIST, even in its revised version, will prove to be less reliable and feasible under the routine clinical conditions of busy primary care practices. Finally, self-report screening tests such as ASSIST may be subject to additional response distortion in primary care settings when patients feel embarrassed, threatened or defensive about their substance use.

Despite the challenges of research on substance abuse screening tests, there are clear benefits to the coordination of instrument development projects such as ASSIST across national boundaries. Most screening tests for drug abuse have been developed in the United States, where the drug use patterns, drug abuse problems and treatment responses differ significantly from other parts of the world. By developing a screening test in a multi-site international study, cross-cultural applicability is addressed at the design stage and feasibility problems are likely to be detected early and solved quickly because of the input available from diverse collaborators. Multi-site studies also have the advantage of expediting the design, evaluation and marketing of a new screening test, to the extent that the collaborating investigators are each capable of introducing the test and continuing its development within their own cultural and linguistic groups.

In 1982, the World Health Organization initiated a program to develop an international screening test for hazardous and harmful alcohol use (Saunders *et al.* 1993). The program led to the creation of the Alcohol Use Disorders Identification Test (AUDIT), which has not only been found to be reliable and valid in numerous evaluation studies (Allen *et al.* 1997), but has also been used widely in primary care and other health settings throughout the world as part of screening and brief intervention programs (Ustun & Sartorius 1995; Babor & Higgins-Biddle 2000). The success of AUDIT raises the question of whether a similar international screening test could be developed for the detection of drug abuse. The results of this study suggest that a screening for drug abuse in the context of a more generic approach that includes alcohol and tobacco products is both reliable and feasible. Further research will be necessary to determine whether other prerequisites to a public health approach to secondary prevention are possible, specifically the collection of valid information free of response bias and the application of effective interventions to people at risk.

NOTE ON AUTHORSHIP

The WHO ASSIST Working Group consists of: Robert Ali[1], Elia Awwad[2], Thomas Babor[3], Fiona Bradley[5], Tecla Butau[6], Michael Farrell[7], Maria Lucia O. S. Formigoni[8], Richard Isralowitz[9], Roseli Boerngen de Lacerda[4], John Marsden[7], Bonnie McRee[3], Maristela Monteiro[10], Hemraj Pal[11], Maritza Rubio-Stipec[12] and Janice Vendetti[3].

[1] Drug and Alcohol Addiction Services Council, Adelaide, Australia; [2] Palestine Red Crescent Society, Palestinian Territories; [3] Department of Community Medicine, University of Connecticut Health Center, Farmington, CT, USA; [4] Department of Pharmacology, Federal University of Parana Curitiba, Brazil; [5] Department of Community Health, Trinity College Dublin, Ireland; [6] Department of Psychiatry, University of Zimbabwe, Harare, Zimbabwe; [7] National Addiction Centre, London, United Kingdom; [8] Universidade Federal de Sao Paulo, Sao Paulo, Brazil; [9] Department of Social Work, Ben Gurion University, Beer Sheva, Israel; [10] Substance Abuse Department, World Health Organization, Geneva, Switzerland; [11] Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India; [12] Behavioral Research Science Institute, University of Puerto Rico, Medical Science Campus, San Juan, Puerto Rico.

ACKNOWLEDGEMENTS

This study was supported in part from funds provided by the World Health Organization, Project on Identification and Management of Psychoactive Substance Problems in Primary Health Care Settings. The authors would like to thank the many helpful staff who contributed to the completion of this study, and the University of Connecticut Alcohol Research Center (NIAAA grant no. 5P50-AA03510-24) for logistical support in the coordination of the project. Drs Ali, Babor, Farrell, Marsden, Monteiro and Rubio-Stipec constituted the Steering Committee that directed the project. Dr Babor was the author responsible for drafting this manuscript. The authors thank Donna Damon for contributions to data management and Deborah Talamini for administrative support. Further information about the WHO ASSIST may be obtained by writing to: Department of Mental Health and Substance Dependence, World Health Organization, 1211 Geneva 27, Switzerland.

REFERENCES

Allen,J. P., Litten,R. Z., Fertig,J. B. & Babor,T. F. (1997) A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, **21**, 613–619.

Babor,T. F. (2002) Is there a need for an international screening test? The Middle East as a case in point. In: *Drug Problems, Cross-Cultural Policy and Program Development*, pp. 165–179. Westport, CT: Auburn House.

Babor,T. F. & Del Boca,F. K. (1992) Just the facts. Enhancing measurement of alcohol consumption using self-report methods. In: *Measuring Alcohol Consumption Psychosocial and Biochemical Methods*, pp. 3–19. New Jersey: Humana Press.

Babor,T. F. & Higgins-Biddle,J. C. (2000) Alcohol screening and brief intervention: dissemination strategies for medical practice and public health. *Addiction*, **95**, 677–686.[PubMed link]

Brown,R. L. & Rounds,L. A. (1995) Conjoint screening questionnaires for alcohol and other drug abuse: criterion validity in a primary care practice. *Wisconsin Medical Journal*, **94**, 135–140.[PubMed link]

MacAndrew,C. (1965) The differentiation of male alcoholic outpatients from non-alcoholic psychiatric outpatients by means of the MMPI. *Quarterly Journal of Studies on Alcohol*, **26**, 238–246.

McPherson,T. L. & Hersch,R. (2000) Brief substance use screening instruments for primary care settings: a review. *Journal of Substance Abuse Treatment*, **18**, 193–202.[PubMed link]

Saunders,J. B., Aasland,O. G., Babor,T. F., de la Fuente,J. R. & Grant,M. (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption. II. *Addiction*, **88**, 791–804.[PubMed link]

Skinner,H. A. (1982) The Drug Abuse Screening Test. *Addictive Behaviors*, **7**, 363–371.[PubMed link]

Sudman,S. & Bradburn,N. M. (1982) *Asking Questions. A Practical Guide to Questionnaire Design.* San Francisco, CA: Jossey-Bass.

United Nations Drug Control Programme (UNDCP) (1997) *World Drug Report.* New York: Oxford University Press.

US Preventive Services Task Force (1998) *Guide to Clinical Preventive Services*, 2nd edn. Washington, DC: DHHS.

Ustun,T. B. & Sartorius,N. (1995) *Mental Illness in General Health Care An International Study.* London: John Wiley & Sons.

Wolff,K., Farrell,M., Marsden,J., Monteiro,M., Ali,R., Welch,S. & Strang,J. (1999) A review of biological indicators of illicit drug use, practical considerations and clinical usefulness. *Addiction*, **94**, 1279–1298.[PubMed link]