RUNNING HEAD: Initial Validation of the ASI–MV

Initial Validation of a Computer-Administered Addiction Severity Index: The ASI–MV

Stephen F. Butler Simon H. Budman

Risa J. Goldman

Innovative Training Systems, Inc, Newton, MA

Frederick L. Newman

Florida International University

Kristen E. Beckley Deborah Trottier

Innovative Training Systems, Inc, Newton, MA

John S. Cacciola

University of Pennsylvania

Manuscript of Published Article: Butler, S.F., Budman, S.H., Goldman, R.J., Newman, F.L., Beckley, K.E., Trottier, D. & Cacciola, J.S. (2001). Initial validation of a computer-administered addiction severity index: The ASI–MV. *Psychology of Addictive Behaviors*, *15*, 4-12.

Abstract

This article introduces the Addiction Severity Index-Multimedia Version (ASI–MV), a computer-administered version of the Addiction Severity Index (ASI). A CD-ROM program was created to simulate a live interviewer. A field trial examined the test-retest reliability, criterion validity and convergent/discriminant validity of the ASI–MV. Participants were 202 clients in treatment at five clinical settings. The ASI–MV demonstrated excellent test-retest reliability on the Composite Scores and Severity Ratings. Criterion validity, tested against the interviewer-administered ASI, yielded good validity for the Composite Scores, but less so for the Severity Ratings. Considerable variability was observed for each of the interviewers, suggesting a lack of interrater reliability among the interviewers. This conclusion was bolstered by the finding that the convergent/discriminant validity of the ASI–MV for both the Composite Scores and Severity Ratings was superior to the standard ASI. The computer-administered ASI is a viable alternative to the expensive and potentially unreliable interviewer administered version.

Initial Validation of a Computer-Administered Addiction Severity Index: The ASI–MV

The Addiction Severity Index (ASI) is a structured clinical interview used to measure the severity of a range of problem areas typically associated with alcohol and drug abuse: Medical Status, Employment Status, Drug Use, Alcohol Use, Legal Status, Family and Social Relationships, and Psychiatric Status (Hendricks, Kaplan, VanLimbeek,& Geerlings, 1989; Hodgins, D.C. & El-Guebaly, 1992; McLellan, Luborsky, Woody, & O'Brien, 1980; McLellan, Luborsky, Cacciola, Griffith, Evans, Barr, & O'Brien, 1985a; McLellan, Kushner, Metzger, Peters, Smith, Grissom, Pettinati, & Argeriou, 1992). This article introduces the Addiction Severity Index-Multimedia Version or ASI–MV. We present the rationale for developing a self-administered, multimedia version of the ASI, a description of the creation of the program and the test-retest reliability, criterion validity and discriminant validity of the ASI–MV.

The Addiction Severity Index: A Standard in the Field. The ASI has become a standard assessment measure in clinical trials and other controlled research involving substance abusers. The Veteran's Administration, many criminal justice agencies, and the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) recommend use of the ASI as part of a comprehensive assessment process, and several states mandate its use. Yet correct implementation of the ASI is very costly to treatment facilities. The ASI is a 45-60 minute interview administered by a trained clinician or interviewer, who must spend another 10 to 20 minutes scoring it (McLellan, Luborsky, Cacciola, Griffith, McGahan, & O'Brien, 1985b).

Interviewer training is expensive as are the costs of refresher workshops to prevent "rater drift."

In response to pressures to use the ASI, some treatment centers simply mandate the ASI, while providing little or no training. The senior authors (SFB and SHB) conducted focus groups on the ASI in May 1999 with 34 substance abuse counselors at a meeting of the National Association of Alcoholism and Drug Abuse Counselors (NAADAC). Counselors who use the ASI reported that training at their clinical sites was spotty at best. Senior counselors often reported that training was reduced from the standard two-day course to merely handing the ASI to new counselors and mandating its use. Such anecdotes are ubiquitous. This lack of training raises serious questions about the quality and relevance of ASI data currently being collected.

The essential subjectivity inherent in the framing of questions, making ratings, and scoring also influences the quality of the data. Dr. McLellan (personal communication, 1997) has noted that interviewers sometimes experience pressure to ensure that clients' ASI results are severe enough to justify admission. Avoidance of such bias is clearly in the interest of the field.

An interactive, multimedia, self-report ASI. A ground-swell of interest in a self-report version of the ASI has resulted in recent efforts to develop a self-report ASI (cf., Rosen, Henson, Finney, & Moos, in press; Block, Mather, & Hallett, 1997; Cacciola, McLellan, Alterman, & Mulvaney, 1998). While these efforts are promising, by definition, pencil and paper self-report assessments cannot address concerns outlined by McLellan, Parikh, Bragg, Cacciola, Fureman, and Incmikoski (1990) regarding illiteracy and establishing a "relationship" with the interviewee.

Multimedia technology may overcome some of the drawbacks of self-report. Concerns about illiteracy can be minimized using audio (spoken text) as well as video qualities of

interactive media. Presentation of interview questions by on-screen "personalities" may approximate a rudimentary "relationship" with the client. Indeed, it is possible that less interpersonal contact offered by the computer may enhance the quality of information obtained from clients. If clients feel less threatened by judgements of an interviewer, they might give more accurate information about substance use and other socially undesirable life circumstances. This has been found to be true in computer surveys of sensitive information such as high-risk sexual activities or injection drug use (e.g., Navaline, Snider, Petro, Tobin, Metzger, Alterman,& Woody, 1994; Turner, Ku, Rogers Lindberg Pleck & Sonenstein, 1998). Based on this rationale, we set out to develop a self-administered, interactive, multimedia version of the ASI.

One problem with a computerized, self-administered ASI is that, without an interviewer, there can be no Interviewer Severity Ratings (ISRs). Severity Ratings for each of the ASI domains comprise one of the two scores traditionally used to summarize client data on the ASI. Unlike the Composite Scores (CSs), which are derived from equations, the ISRs are the subjective ratings of interviewers. Although the ASI's authors caution against use of the severity ratings (McLellan, et al., 1990), they remain a popular summary statistic in clinical settings. To address this problem, we derived regression equations that result in Predicted Severity Ratings (PSRs) for each ASI domain using several large datasets of research-based severity ratings (Butler, Newman, Cacciola, Frank, Budman, McLellan, Ford, Blaine, Gastfriend, Moras, & Salloum, 1998). These equations permit the calculation of Severity Ratings that achieve reliability coefficients with expert ASI interviewers (IntraClass Correlations from .64 to .96) that are as good as or better than specially trained human interviewers. Thus, the PSRs permit the

computer to estimate ISRs. Furthermore, since standard ASI training does not include formal tests of interrater reliability, these equations may provide superior estimates of experts' ratings and ratings that do not drift or are otherwise biased.

Conclusion. A client self-administered, computerized ASI would represent a rational distribution of resources for financially hard-pressed agencies. It would free clinical staff to attend to direct clinical matters. It is conceivable that a self-administered computerized version of the ASI, such as the ASI–MV, would come to reflect greater standardization of the ASI assessment process, uninfluenced by variability in training procedures, rater "drift," and pressures to meet certain outcome or admission criteria.

The primary objective of this research was the development of a self-administered computer version of the ASI, the ASI–MV, and to conduct evaluations of its test-retest reliability (Study 1), criterion validity (Study 2) and discriminant validity (Study 3). To our knowledge, this is the first attempt to rigorously test a self-administered, computerized version of the ASI.

Method

Development of the ASI–MV

The initial step in the development of the ASI–MV was to create a process flowchart reflecting the logic associated with interviewer administration of the ASI and the rules specified in the ASI Manual (McLellan et al., 1990). The flowchart characterizes the decision tree that interviewers explicitly and implicitly follow to organize the interview. Two ASI experts, Drs. John Cacciola and Sabrina Ford, created a detailed flow chart. The flow chart formed the basis of a "script" for the program, used by actors to create the program's audio and video. Each ASI

domain is introduced by a different virtual "interviewer," corresponding to the theme of the domain. For instance, a virtual physician in her office and white coat asks the medical items, while a virtual drug counselor asks the drug and alcohol questions.

Field Trial Facilities

The field trial was conducted at five substance abuse treatment centers in New England. Sites were selected to achieve a multi-state sample of clients with a range of characteristics and from a variety of treatment settings. Two sites were located in New Hampshire, a prison diversion facility and a residential facility. The New York site was both a day treatment and residential facility. The sites in Massachusetts and Rhode Island were residential facilities.

ASI Interviewers and Training Procedures

Seven staff members were referred by the participating sites to undergo ASI training and administer the ASI for the research project. The standard, two-day ASI training was conducted by DeltaMetrics, Inc. in order to ensure that official ASI training procedures were utilized. Dr. McLellan is the Scientific Director of DeltaMetrics, and this company is the primary source of ASI training for research and clinical purposes. The training covered the goals and objectives of the ASI, coding and rating procedures, a coding quiz, scoring of video vignettes, and skill building exercises including roleplays. Two months following the training, a quiz on ASI knowledge was administered and scored by DeltaMetrics, and feedback given to the interviewers. Five months after training, a one-day booster session was conducted by DeltaMetrics. Finally, a scored ASI from each rater was sent to DeltaMetrics for review, comments and feedback. These procedures reflect state-of-the-art ASI training.

Participants

Two-hundred-two (202) clients between the ages of 18 and 75 participated in the study. Inclusion criteria for subject participation included: primary diagnosis of substance abuse or dependence, in treatment for at least three days, deemed stable enough to complete study procedures, and willingness to sign informed consent. Stabilization of the client was a subjective decision made by the treatment staff. Subjects were compensated for participation.

Comparison Measures

Comparison measures, administered to participants in the convergent/discriminant validity study, were selected to be comparable to those used by McLellan et al. (1985a). The measures included: (1) The Physical Component Summary of the SF-12 (Ware, Kosinski, & Keller, 1995), a 12-item measure of four physical health concepts including physical functioning, role limitations due to physical problems, bodily pain and general health, was used to test the medical domain of the ASI. (2) The Work subscale of the Social Adjustment Scale-Self Report (SAS-SR; Weissman & Bothwell, 1976) was used for comparison with the ASI employment scores. The SAS-SR is a 54-item self-report scale for measuring instrumental or expressive role performance in major areas of functioning. The Work subscale combines reports of work outside the home, work in the home (homemaker) and work as a student. (3) The ASI Alcohol domain was compared with the Michigan Alcoholism Screening Test (MAST; Selzer, 1971). The MAST is a 25-item questionnaire which screens for and assesses seriousness of alcohol abuse. (4) The ASI Drug domain was compared with the Drug Abuse Screening Test (DAST, Skinner, 1982). Modeled after the MAST, the DAST is a 20-item assessment of drug abuse and dependence. (5)

The Legal domain comparison measure was the Antisocial Behavior Checklist (ABC, Zucker & Fitzgerald, l996), a 46-item measure which taps clients' histories of antisocial behaviors such as running away from home, defaulting on debt, resisting arrest, and so forth. The ABC differentiates individuals with long histories of criminal activity from others (Zucker, Noll, Ham, Fitzgerald & Sullivan, l994). (6) The Family/Social domain comparison measure was the Extended Family subscale of the Social Adjustment Scale-Self Report (SAS-SR, Weissman & Bothwell, 1976). This subscale taps concepts around family attachment and level of conflict within the family. Finally, (7) the ASI Psychiatric domain was compared with the Hopkins Symptom Checklist-90 Item (HSCL-90; Derogatis, Lipman, & Rickels, 1974), a popular self-report measure of a wide array of psychiatric symptoms. The global severity score of the HSCL-90 has good psychometric properties and is widely accepted as a measure of psychiatric distress.

Design of the Field Trial and Procedures

The overall design called for three studies: Study 1 was an examination of the test-retest reliability of the ASI–MV, Study 2 was an assessment of criterion validity, and Study 3 examined convergent/discriminant validity. Inclusion criteria were the same across all studies.

Study 1 Procedures to Examine Test-retest Reliability

Sixty participants completed the ASI–MV at two times separated by three to five days. This procedure was similar to that of McLellan and colleagues (1985a), who argued that a three-day separation would be enough to reduce the likelihood of participants' simply repeating answers from memory and short enough to reduce the possibility of genuine changes in the clients' situations. After a brief introduction, all clients completed the ASI–MV without

assistance. Three to five days later (depending on participants' schedules), a return visit was scheduled for the participant to complete the ASI–MV a second time.

Study 2 Procedures to Examine Criterion Validity

Criterion Validity was evaluated by comparing the ASI–MV to a "gold standard," in this case, the standard, interviewer-administered ASI. One-hundred forty-two participants were administered both the ASI–MV or the standard, interview ASI three to five days apart. Order of administration was counter-balanced in order to minimize any order effect.

Study 3 Discriminant Validity

One-hundred ten clients participated in this study. Participants completed the ASI–MV followed by the comparison measures listed above. Some participants (N = 60) also served in the test-retest study and some (N = 50) also served in the criterion validity study described above. Since those who served in the criterion study also were administered the standard, interviewer version of the ASI, these participants' data were used to examine the discriminant validity of the standard version of the ASI.

Results

Client participants' characteristics

Two-hundred and two (202) clients in substance abuse treatment served as participants in the study and were exposed to the ASI–MV. Participants for the reliability and validity studies overlapped, so that some participants' data contributed to two studies. All participants were exposed to the ASI–MV program.

Demographic data were obtained from answers to the ASI–MV. Characteristics of the

overall sample and by study are summarized in Table 1. As can be seen, participants' average age was 34.4, with a standard deviation of 9.3 years, and a range from 18 to 73 years. The sample was comprised mostly of men (62.4%) and were mostly white (64.4%), thus a bit more than a third of the sample (35.7%) were minority participants. A national survey of substance abuse treatment in the US (SAMHSA, 1999) reports that about 32% of substance abusers in treatment nationally were female and about 57% white, 25% African-American, 14% Hispanic, and 3% other. Thus, the present sample generally reflects the ethnic and gender distributions found nationally for substance abusers in treatment. Our sample slightly under-samples Hispanic persons in treatment. Most participants used alcohol (85%), cocaine (64%) and marijuana (72%). The largest percentage (42%) reported alcohol and drug use as their primary problem, followed by alcohol only (20%) and polydrug use (12%). Finally, mean education was 12 years, however, nearly 30% of the sample had less than a high school education.

Use of the ASI–MV by clients in substance abuse treatment

Early discussions with counselors raised concerns that substance abusers would be overwhelmed by the effort required of the computer program. Indeed, we found the opposite, at least with the 202 volunteers who signed informed consent to participate in the study. No person who began the study rejected the program or failed to complete it through to the end. In fact, there were spontaneous reports suggesting that the participants were very positive about the experience and found it "easier" than "going over all this with a person."

Another concern was the length of time it would take clients to complete the program. The ASI interview is said to average 45 minutes to an hour, although a person with a complicated

history can take much longer. If the average ASI–MV session was much longer than an hour, enthusiasm for the program would likely be reduced. In fact, mean time for completion of the ASI–MV was 42.6 minutes (SD = 16.3 minutes), with a median of 38 minutes, and a range of 19 to 123 minutes (two hours and three minutes). Only three of 202 Ss took more than 90 minutes and 13.4% took more than 60 minutes. Thus, the length of time to complete the program appears to be acceptable.

We next looked at the effect of age, race (white versus minority) and gender on time to complete the ASI–MV. A moderate, positive correlation was found for age (.33), suggesting that age accounted for about 11% of the variance observed in the completion times. There was no significant difference for gender, with men taking an average of 43 minutes, and women averaging 42 minutes. However, whites took an average of 39 minutes, and minority clients averaged 49 minutes ($t = 4.3$, $df = 199$, $p < .001$). Despite this difference, the means are within acceptable limits. Finally, we looked for a possible effect of problem severity (as measured by the CSs and PSRs) on ASI–MV completion time. No clear association was observed, with the strongest correlation being the CS for employment, which correlated .28, accounting for about 8% of the variance in completion time.

Study 1: Test-retest reliability for Composite scores and Severity ratings

Test-retest reliability correlations are presented in Table 2 for the Composite scores (CSs) and Predicted Severity Ratings (PSRs), the computer-generated estimation of the ISR. Two Ss (3.3%) who completed the Time 1 test-retest session did not return for their Time 2 session. The correlations for the CSs are quite good, ranging from a low of .68 for the legal domain, to a high

of .95 for the employment domain. For the PSRs, the test-retest reliability was acceptable, ranging from a low of .62 for the legal domain to a high of .84 for psychiatric domain. Clearly the ASI–MV test-retest reliability for these two summary scores appears satisfactory.

Study 2: Criterion validity for Composite scores and Severity ratings

The concept of criterion validity in the present context is complex. Criterion validity contrasts the measure to be validated against some criterion or "gold standard" (Anastasi, 1976; Nunnally, 1978). Typically, when another measure is used as the criterion, a pearson correlation is used to establish evidence of criterion validity. For the ASI–MV, the "gold standard" is the interviewer-administered ASI. The correlations obtained between Composite Scores for the ASI–MV and the standard administration of the ASI were positive, highly significant ($p < .001$), and generally of reasonable magnitude. The correlations for the Composite Scores for each ASI domain were as follows: Medical domain $r = .71$, Employment domain $r = .90$, Alcohol domain $r = .67$, Drug domain $r = .62$, Legal domain $r = .59$, Family/Social domain $r = .37$, and Psychiatric domain $r = .76$. The correspondence between the ASI–MV PSRs and the Interviewers' Severity Ratings was less clear. For these correlations, Medical domain $r = .50$ ($p < .001$), Employment domain $r = .28$ ($p = .002$), Alcohol domain $r = .19$ ($p = .04$), Drug domain $r = .12$ (NS), Legal domain $r = .46$ ($p < .001$), Family/Social domain $r = -.06$ (NS), and Psychiatric domain $r = .43$ ($p < .001$). Note that during the criterion test, 7.7% (N = 11) Ss did not complete the Time 2 administration.

Another way of thinking about the ASI–MV's relationship to the standard, interview version views the ASI–MV as a new "interviewer," whose interrater reliability must be

established with interviewers who have undergone the standard ASI interviewer training program. From this perspective, the IntraClass Correlation (ICC) may be the most appropriate statistic. The ICC statistic is recommended when the variables to be correlated belong to a common class, meaning that the variables share both their metric and variance (McGraw & Wong, 1996), as opposed to the pearson r, which is used when measures have different scales (e.g., weight and height). Thus, the ICC is typically used as an index of interrater reliability to examine the correspondence of the ratings of two or more raters judging the same items. Interpretations of the magnitude of ICCs generally assume that values greater than .80 represent perfect agreement, .61 to .80 is substantial, .41 to .60 is moderate, and .21 to .40 is fair reliability (Landis & Koch, 1977).

The two right-hand columns of Table 2 present correspondence of the ASI–MV and the interviews for Composite and Severity Scores. Composite Scores for all raters is quite good, ranging from a high of .95 for the Employment domain to a low of .54 for the Family/Social domain, with all but the Family/Social Domain having greater than .70 correspondence. While this correspondence is clearly acceptable, we noted considerable variation among the seven different interviewers. For instance, ICCs for the Legal domain ranged from .36 for one interviewer to .89 for another, and ICCs for the Family/Social domain ranged from a .38 for one interviewer to .94 for another.

This trend was even more marked for the Severity Rating ICCs. As with the pearson correlations, much less correspondence was observed, the ICCs ranged from a high of .64 for the Medical domain to a low of -.12 for the Family Domain (Table 2). However, the difference in

ICCs across interviewers were quite extreme. For example the Family/Social domain ranged from a very poor -.37 for one interviewer to .71 for another. These discrepancies in agreement across interviewers are hard to interpret since the standard ASI training does not establish interrater reliability of the raters themselves. Without confirmation that the ASI interviewers were reliable with each other, and given the wide range of ICCs obtained for each interviewer, it is impossible to decisively evaluate the ASI–MV with regard to the Severity Rating.

Disagreement between the computer-generated and interviewer Severity Ratings can be further explored by calculating PSRs on the interview data. If interviewers are eliciting different information from clients than the ASI–MV, and their judgments are the result of this different information, then poor agreement should also be observed when comparing the computer PSRs with PSRs calculated using the interview data. If, on the other hand, the ASI–MV and interview PSRs correspond better, this would at least be suggestive that the discrepancies observed between the computer- and interviewer-generated Severity Ratings are the result of the judgments made by the interviewers. Indeed, ICCs calculated between the ASI–MV PSRs and the PSRs calculated on the interview data suggest much better agreement. These ICCs are as follows: for the Medical domain, .76; Employment domain, .64; Alcohol domain, .69; Drug domain, .49; Legal domain, .84; Family and Social domain, .74; and Psychiatric domain, .81. Compared with the ICCs for all raters reported in Table 2, these ICCs are of generally acceptable magnitude. This clear improvement over the ICCs obtained with the interviewers' Severity Ratings suggests that respondents were not reporting drastically different basic information to the computer or the interviewer. Rather, it supports the hypothesis that the observed discrepancy in

severity rating agreement is the result of problems in the interviewers' judgment, the very subjectivity with which McLellan and colleagues (1990) have been concerned.

Study 3: Discriminant validity for Composite scores and Severity ratings

We attempted to replicate the convergent/discriminant validity test conducted by McLellan et al. (1985). This conservative assessment of validity, discussed by Campbell and Fiske (1959), requires three conditions be met. First, each ASI–MV domain score (Composite or Severity Score) should be correlated in the appropriate direction with its designated comparison test. Second, each domain score should be more highly correlated with its designated comparison than with any of the other tests. Finally, a comparison test should be more highly correlated with its paired ASI domain score than with any of the other ASI domain scores.

Reasonable variation was obtained for the comparison measures. Sample means and standard deviations for the comparison measures are as follows: Physical Component Summary of the SF-12 mean was 45.3 (SD = 9.2), the SAS-SR Work subscale mean was 3.5 (SD = 1.75), the MAST mean was 31.7 (SD = 13.7), the DAST mean was 12.4 (SD = 6.0), the ABC mean was 37.6 (SD = 22.1), the SAS-SR Extended Family subscale mean was 2.3 (SD = .8), and finally, the HSCL-90 mean for average scores was 2.0 (SD = .6).

The correlation matrix for the Composite Scores is presented in Tables 3 for the ASI–MV and the standard interviewer version. The boxed values are the correlations among the comparison tests with their paired ASI domain score[1]. As can be seen in Table 3, there was general evidence of good discriminant validity for the ASI–MV. That is, in most cases, the ASI–MV Composite Score correlated higher with the comparison measures (on-diagonal

correlations) than with the other comparison measures (off-diagonal correlations). Three Composite Scores (Employment, Legal, and Family/Social) had higher correlations with the HSCL-90. Similar overlap of the correlations was observed by McLellan et al. (1985a). Some of the obtained on-diagonal correlations were not high. For instance, the Employment Composite Score correlated only .17 with the Social Adjustment Scale-total work score and the Legal Composite Score correlated only .18 with the Antisocial Behavior Checklist (ABC). These low correlations might reflect different foci of the measures. For instance, the SAS emphasizes emotional and interpersonal work adjustment, whereas the ASI emphasizes current employment and work history. Likewise, the ABC emphasizes personality issues associated with antisocial personality, whereas the ASI Legal domain focuses on current and past legal history and problems. Nevertheless, the evidence for discriminant validity remains strong. The average on-diagonal correlation (.37) was significantly higher than the average off-diagonal correlation (.17) for the ASI–MV ($\underline{t} = 3.92$; $\underline{df} = 47$, $\underline{p} < .001$). Interestingly, the convergent/discriminant validity of the Composite Scores for the interviewer-administered ASI was less in evidence. Although the basic pattern of relationships was similar for both versions of the ASI, the magnitude of the correlations with the comparison measures was generally lower for the interviewer-administered ASI than the ASI–MV. Furthermore, the average on-diagonal correlation (.28) for the interviewer version was not significantly different from the average off-diagonal correlation (.17; $\underline{t} = 1.82$; $\underline{df} = 47$, NS). This finding differs from the results of the discriminant validity test reported by McLellan and colleagues (1985a).

Similar results were obtained for the Severity ratings as presented in Table 4. Again,

there was general support for the discriminant validity of the ASI–MV Severity ratings, with the average on-diagonal correlation (.44) significantly higher than the average off-diagonal correlation (.18; $t = 5.85$; $df = 47$, $p < .001$). And again, the average correlations for the interviewer-generated severity ratings were not significantly different (on-diagonal mean = .24; off-diagonal mean = .15, $t = 1.78$; $df = 47$, NS).

These discriminant validity data illuminate certain concerns raised by the evaluation of criterion validity discussed above. Consider, for instance, those domains for which correspondence was not particularly good between the ASI–MV and the criterion, namely the standard interviewer-administered ASI. For the Composite Scores, the lowest correlation (.37) was for the Family/Social domain. Examination of the convergent/discriminant validity matrix for this domain reveals much greater convergent validity for the ASI–MV (.43) than for the interviewer version (.27). For the Severity ratings this pattern was even stronger. For those domains where the correspondence with the interview was less than .25, in every case the ASI–MV's convergent/discriminant validity was superior to the interview. For example, Severity Ratings on the two versions for the Alcohol domain correlated only .19, but the ASI–MV correlated .43 with the MAST while the interviewers' ratings correlated only .06. For the Drug domain, the correlation between the two versions of the ASI was only .12, but the ASI–MV correlated .59 with the DAST, while the interviewers' ratings correlated only .25. Again, the Family/Social domain achieved a negative correspondence between the two versions (-.06), but the ASI–MV's correlation with the SAS comparison measure was a little larger (.37) than achieved by the interviewers (.31). These data suggest that where the ASI–MV did not

"agree" with interviewers, its convergent validity was at least as good as and generally better than the interviewer-administered ASI.

Discussion

This project endeavored to develop and test a multimedia computer program to simulate an interviewer administering the ASI. The ASI–MV models the logic implied in the ASI interview and uses empirically derived formulas to predict interviewers' severity ratings. Once developed, the ASI–MV was subjected to tests of test-retest reliability, criterion validity and convergent/discriminant validity.

Overall, the ASI–MV performed quite well. Test-retest reliability was excellent, suggesting that the ASI–MV estimates of the Composite Scores and Severity Ratings are stable. Criterion validity was tested using the standard, interviewer-administered ASI as the criterion. Results suggested that criterion validity for the Composite Scores was generally good overall, although the Severity Ratings yielded less concordance. For both the Composite Scores and Severity Ratings, however, considerable variability was observed for each of the interviewers. That is, no domain score had uniformly poor correspondence for all interviewers. This variability was interesting in light of the fact that the PSR equations are invariant.

The interviewers in this study were trained in accordance with the standard training offered for the ASI. This training, although comprehensive and recommended for both research and clinical purposes, does not directly assess interrater reliability. Without a determination of the interrater reliability of the ASI interviewers, it is difficult to interpret the variable correlations observed with the ASI–MV, especially for the Severity Ratings. Findings of better concordance

using PSRs calculated on the interview data along with the finding that ASI–MV PSRs may have better convergent/discriminant validity than the standard ISRs bolster the argument that the problem may lie more with the interviewers and less with the ASI–MV.

The convergent/discriminant validity analysis provided strong support for the ASI–MV as a measure at least as valid and perhaps more so than the ASI interview. It is interesting to note that the correlations obtained for the interview were in some cases less than those obtained by McLellan and colleagues (1985a) in their discriminant analysis. It is not entirely clear why we were unable to replicate McLellan's findings. Several possible explanations present themselves. One possibility is that we compared correlations of the ASI–MV and the interview ASI with pencil and paper self-report measures. Since the ASI–MV is more of a self-report format than is the interview ASI, the apparent superior performance of the ASI–MV may be due to shared method. This explanation still does not explain our inability to replicate the magnitude of McLellan's correlations, since the original discriminant validity test also used pencil and paper self-report measures (McLellan et al., 1985a).

Another, perhaps more likely possibility may revolve around differences in the training and supervision of the interviewers in the two studies. Although we used recommended training procedures, it is likely that the training and supervision of interviewers was not as intensive as in the original validation attempt. Others (e.g., Hodgins and El-Guebaly, 1992; Alterman, Brown, Zaballero & McKay, 1994) have reported similar difficulty replicating the excellent test characteristics reported by McLellan et al. (1985a), supporting the notion that interviewers in the original ASI validation study may have been particularly well-trained and monitored, beyond

what is generally available in the typical clinical setting.

It is important finally to address the practicality of using a computer to conduct the ASI in a real-world clinical setting. The ASI–MV was administered to 202 clients in treatment and was not rejected by any clients. Clients understood how the program worked and got through it with little difficulty. The time to administer the ASI–MV appeared to be comparable to administration times obtained with the interview. However, by using the ASI–MV, clinician time is reduced to a minimum. Because the computer automatically scores the ASI and prints a clinical report, it should further reduce staff time. The ASI–MV appears to be practical for use in a clinical setting.

Several products currently exist that allow clinicians to hand-enter ASI interview data into a computer. These programs streamline scoring and reporting. But these programs still require substantial staff time to administer the interview and input the data, and the ASI results are still subject to bias and inconsistency. The ASI–MV should be very cost-effective, even for resource-poor clinical settings. We estimate that a facility currently using the ASI could buy a new computer and the program for less than the cost of 15 traditional administrations, and continue to save nearly 80% for each subsequent ASI–MV administered.

Most of the limitations to the ASI–MV are similar to those for the standard version of the ASI. Clinicians and researchers need to determine if a client's current mental status is appropriate for an extended session on a computer. If the client is disorganized, highly agitated, intoxicated, actively in withdrawal, of limited cognitive capacity or otherwise not appropriate to take the ASI–MV, alternative assessment approaches will be necessary. The ASI–MV, like the

standard ASI, is not intended as an assessment for adolescents (NIDA, 1993).

None of the volunteer participants in this study rejected the computer administration, and it should be recalled that about 30% of the sample had less than a high school education. However, it is clearly possible in clinical settings that some individuals might simply reject the computer based on a pre-existing bias against computers. Our experience discussing the ASI–MV with counselors suggests that counselors may underestimate their clients' willingness to interact with a computer. A positive and matter-of-fact attitude on the part of the staff person who introduces the ASI–MV should minimize rejection rates by clients. Nevertheless, the numbers of non-research clients who will reject the computer remains a question.

The ASI–MV is a computer-administered version of the ASI that requires minimal staff time and has client-completion time that is comparable to the standard interview. The ASI–MV demonstrates acceptable test-retest reliability, good criterion validity and acceptable convergent/discriminant validity, especially in comparison to the standard ASI. The ASI–MV should be as acceptable for most clinical and research purposes as the standard ASI. The present data suggest further that the computer version may not only provide a more economical assessment than the interview, it may generate more reliable and valid data than a standard ASI conducted by trained (and certainly untrained) interviewers.

References

Alterman, A.I., Brown, L.S., Zaballero, A., & McKay, J.R. (1994). Interviewer severity ratings and composite scores of the ASI: a further look. Drug and Alcohol Dependence, 34, 201-209.

Anastasi, A. (1976). Psychological testing. 4th ed. New York: Macmillan.

Block, P., Mather, D. & Hallett, R. (1997, November). Development of a self-report version of the Addiction Severity Index. Paper presented to the Association for the Advancement of Behavior Therapy. Miami, FL.

Butler, S.F., Newman, F.L., Cacciola, J.S., Frank, A., Budman, S.H., McLellan, A.T., Ford, S., Blaine, J., Gastfriend, D., Moras, K., & Salloum, I.M. (1998). Predicting ASI interviewer severity ratings for a computer-administered Addiction Severity Index. Psychological Assessment, 10, 399-407.

Campbell, D.T. & Fiske, D. (1959). Convergent and discriminant validity by the multi-trait multi-method matrix. Psychological Bulletin, 56, 81-105.

Cacciola, J.S., McLellan, A.T., Alterman, A.I., & Mulvaney, F.D. (1998). A comparison of a self-administered ASI with the standard ASI interview. In L.S. Harris (Ed.) Problems of Drug Dependence, 1997: Proceedings of the 59th Annual Scientific Meeting, the College on Problems of Drug Dependence, Inc. NIDA Research Monograph 178. NIH Publication No. 98-4305, p. 245.

Derogatis, L., Lipman, R., and Rickels, K. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. Behavioral Sciences, 19, 1-16.

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed). Wiley & Sons, N.Y.

Hendricks, V.M., Kaplan, C.D., VanLimbeek, J., & Geerlings, P. (1989). The Addiction Severity Index: Reliability and validity in a Dutch addict population. Journal of Substance Abuse Treatment, 6, 133-141.

Hodgins, D.C. & El-Guebaly, N. (1992). More data on the Addiction Severity Index: Reliability and validity with the mentally ill substance abuser. Journal of Nervous and Mental Disease, 180, 197-201.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.

McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some Intraclass Correlation Coefficients. Psychological Methods, 1, 30-46.

McLellan, A.T., Luborsky, L., Cacciola, J., Griffith, J.E., Evans, F., Barr, H.L., & O'Brien, C.P. (1985a). New data from the Addiction Severity Index: Reliability and validity in three centers. Journal of Nervous and Mental Disease, 173, 412-423.

McLellan, A.T., Luborsky, L.,Cacciola, J., Griffith, J.E., McGahan, P., & O'Brien, C.P. (1985b). Guide to the Addiction Severity Index: Background administration and field testing results. U.S. Department of Health and Human Services, Rockville, MD.

McLellan, A.T., Luborsky, L., Woody, G. E., & O'Brien, C.P. (1980). An improved evaluation instrument for substance abuse patients: The Addiction Severity Index. Journal of Nervous and Mental Disease, 168, 26-33.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., and

Argeriou, M. (1992). The fifth edition of the Addiction Severity Index. Journal of Substance Abuse Treatment, 9,199-213.

McLellan, A.T., Parikh, G., Bragg, A. Cacciola, J., Fureman, B., Incmikoski, R. (1990). Addiction Severity Index, Fifth Edition with preface: A guide to training and supervising the ASI Interviews based on the past ten years. Unpublished manuscript, University of Pennsylvania, Center for Studies of Addiction.

National Institute on Drug Abuse (NIDA). (1993). Assessing client needs using the ASI: A handbook for program administrators. NIH pub No. 93-3619, Rockville, MD.

Navaline, H.A., Snider, E.C., Petro, C.J., Tobin, D., Metzger, D. Alterman, A.I., & Woody, G.E. (1994). An automated version of the risk assessment battery (RAB): Enhancing the assessment of risk behaviors. AIDS Research and Human Retroviruses, 10, S281-S283.

Nunnally, J. (1978). Psychometric theory. 2nd Ed. New York: McGraw-Hill.

Rosen, C.S., Henson, B.R., Finney, J.W., & Moos, R.H. (In press). Consistency of self-administered and interview-based Addiction Severity Index composite scores. Addiction.

SAMHSA (1999). Uniform facility data set (UFDS): 1997 data on substance abuse treatment facilities. US Dept of Health & Human Services. Document available from http://www.samsha.gov.

Selzer, M.D. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. American Journal of Psychiatry, 127, 1653-1658.

Skinner, H.A. (1982). The drug abuse screening test. Addictive Behaviors, 7, 363-371.

Turner, CF. Ku L., Rogers S.M., Lindberg L.D., Pleck J.H., & Sonenstein, F.L. (1998).

Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. Science, 280, 867-873.

Ware, J.E., Kosinski, M., & Keller, S.D. (1995). SF-12: How to score the SF-12 Physical and Mental Health Summary Scales. The Health Institute, Boston, MA.

Weissman, M.M. & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. Archives of General Psychiatry, 33, 1111-1115.

Zucker, R. A. & Fitzgerald, H.E. (l996). The Antisocial Behavior Checklist (Version 2). Ann Arbor: Department of Psychiatry and Alcohol Research Center , University of Michigan.

Zucker, R. A., Noll, R. B, Ham, H., Fitzgerald H. E., & Sullivan L. S. (l994). Assessing antisociality with the Antisocial Behavior Checklist: Reliability and validity studies. East Lansing (Michigan State University) and Ann Arbor (University of Michigan), Michigan. MSU-UM Longitudinal Study.

Author Note

This project was supported, in part, by an SBIR grant #2R44DA09938 from the National Institute on Drug Abuse to the first author.

We owe considerable debt to Mr. John Ahman of Marathon, Inc., Providence, RI and Dr. Bruce Maslack of St. Peters Addiction Recovery Centers, Albany, NY. We also wish to acknowledge Dr. Tom McLellan for encouraging us to pursue the idea of a computer-administered ASI.

Correspondence concerning this article should be addressed to Stephen F. Butler, Innovative Training Systems, Inc. 199 Wells Ave., Suite 108, Newton, MA 02459. Electronic mail may be sent via Internet to sfbutler@aol.com.

Footnote

[1]Significance levels are not presented for the correlations in Tables 3 and 4. We were more interested in the magnitude of the correlations than their significance. This is especially true for comparisons for the discriminant validity of the ASI–MV with the interviewer-administered ASI, since the different Ns have different power for establishing significance. Thus, for the ASI–MV, the N of 109 requires only an r value of .26 to reach significance, while the N of 50 for the interviewer-administered ASI requires an r value of .38 to reach significance.

Table 1 Demographic Characteristics of Participants in Overall Sample and by Study

Characteristic	Overall Sample N = 202[1]	Test-Retest Reliability N= 60	Criterion Validity N=142	Discriminant Validity N=110
Age (in years)				
Mean (SD)	34.4 (9.3)	36.3 (8.8)	33.5 (9.5)	35.3 (9.5)
Range	18 – 73	19 - 58	18 – 73	18 – 73
Gender				
Male	62.4%	70.0%	59.2%	67.3%
Female	37.6%	30.0%	40.8%	32.7%
Race				
White	64.4%	66.7%	63.4%	67.0%
African-American	22.8%	15.0%	26.1%	20.2%
Hispanic	9.9%	13.3%	8.4%	9.2%
Native-American	1.5%	1.7%	1.4%	1.8%
Other	1.5%	3.3%	.7%	1.8%
Marital Status				
married/live with SO	14.9%	15.0%	14.8%	14.7%
single	57.4%	51.7%	59.9%	54.1%
divorced/separated	25.2%	31.7%	22.5%	28.4%
widowed	2.5%	1.7%	2.8%	2.8%
Education (in years)				
Mean (SD)	12.0 (2.8)	12.6 (2.3)	11.7 (2.0)	12.0 (2.3)
Range	6 - 20	8 - 20	6 - 20	6 - 20
% < 9th grade	4.5%	1.7%	5.6%	5.5%
% < high school	29.2%	20.0%	33.1%	26.6%
Court ordered treatment	37.8%	39.0%	37.3%	40.7%
Currently on parole or probation	51.7%	51.7%	51.8%	54.1%
Primary Substance of Abuse				
Alcohol Only	20.4%	25.5%	18.4%	20.4%
Heroin	7.7%	5.5%	8.5%	5.8%
Cocaine	11.7%	3.6%	14.9%	6.8%
Cannabis	2.6%	1.8%	2.8%	1.9%
Alcohol & Drugs	45.4%	49.1%	44.0%	53.4%
Polydrug	12.2%	14.5%	11.3%	11.7%

[1]Since studies all Ss participated in more than one "study," the Ns do not add.

Table 2 Test-retest Pearson Correlations (Test-Retest Reliability) for ASI Composite scores and Predicted Severity Ratings (PSRs) and IntraClass Correlations (Criterion Validity) for ASI Composite Scores and Severity Ratings

	Composite Scores	Predicted Severity Scores	Composite Scores ASI–MV & Interview	Severity Scores ASI–MV & Interview
	Test-Retest Pearson Correlations N=58	Test-Retest Pearson Correlations N=58	Criterion IntraClass Correlation N=131	Criterion IntraClass Correlation N=127
Medical Domain	.80	.80	.83	.64
Employment Domain	.95	.95	.95	.40
Alcohol Domain	.86	.86	.80	.26
Drug Domain	.75	.75	.77	.23
Legal Domain	.68	.68	.73	.62
Family/Social Domain	.80	.80	.54	-.12
Psychiatric Domain	.81	.81	.86	.58

Table 3 Correlation Matrix of Composite Scores with Comparison Measures for ASI–MV and Interview

	SF-12 Physical[1]	SAS[2]-Work Total Score	MAST[3]	DAST[4]	ABC[5]	SAS– Extended Family	HSCL-90
ASI–MV Medical Composite Score	-.34	.02	.24	.15	.11	.29	.36
Interview Medical Composite Score	-.39	.00	.02	-.01	-.30	.23	.27
ASI–MV Employment Composite Score	.03	.17	.04	.13	.06	.04	-.07
Interview Employment Composite Score	-.02	.09	-.04	.22	-.16	.10	.05
ASI–MV Alcohol Composite Score	-.04	.14	.39	.20	.13	.21	.35
Interview Alcohol Composite Score	-.27	.11	.22	.12	-.24	.27	.38
ASI–MV Drug Composite Score	.07	.31	.14	.46	.31	.29	.32
Interview Drug Composite Score	-.22	.20	.08	.26	-.05	.31	.42
ASI–MV Legal Composite Score	.10	-.06	-.05	.07	.18	.12	.22
Interview Legal Composite Score	-.03	.04	-.31	.28	-.03	.22	.08
ASI–MV Family/Social Composite Score	.01	.07	.17	.24	.33	.43	.50
Interview Family/Social Composite Score	.03	-.08	.05	.11	.07	.27	.25
ASI–MV Psychiatric Composite Score	-.03	.14	.26	.18	.14	.39	.65
Interview Psychiatric Composite Score	-.12	.27	.19	.27	.16	.59	.70

N for ASI–MV = 109; N for ASI Interview = 50

[1]Short Form-12 Physical Components Scale (Ware et al., 1995), Note: Higher scores on this measure indicate less severity; [2]Social Adjustment Scale (SAS-SR; Weissman & Bothwell, 1976); [3]Michigan Alcoholism Screening Test (MAST; Selzer, 1975); [4]Drug Abuse Screening Test (DAST; Skinner, 1982); [5]Antisocial Behavior Checklist (ABC; Zucker et al., 1994).

Table 4 Correlation Matrix of Severity Ratings with Comparison Measures for ASI–MV and Interview

	SF-12 Physical[1]	SAS[2]-Work Total Score	MAST[3]	DAST[4]	ABC[5]	SAS–Extended Family	HSCL-90
ASI–MV Medical Severity Rating	-.42	-.01	.25	.17	.20	.28	.31
Interview Medical Severity Rating	-.31	.06	-.06	.06	-.23	.26	.32
ASI–MV Employment Severity Rating	-.16	.06	.20	.15	.24	.32	.40
Interview Employment Severity Rating	.04	.04	-.07	.20	-.08	.38	.42
ASI–MV Alcohol Severity Rating	-.15	-.02	.43	.06	.10	.18	.15
Interview Alcohol Severity Rating	-.03	-.11	.06	-.04	-.32	.14	.19
ASI–MV Drug Severity Rating	.03	.11	-.04	.59	.31	.17	.10
Interview Drug Severity Rating	-.08	.00	.02	.25	.03	.26	.24
ASI–MV Legal Severity Rating	.14	-.10	.09	.19	.27	.17	.12
Interview Legal Severity Rating	.01	-.12	-.26	.16	-.19	.18	.09
ASI–MV Family/Social Severity Rating	-.02	.00	.13	.39	.32	.37	.43
Interview Family/Social Severity Rating	.07	-.08	.05	-.02	.04	.31	.35
ASI–MV Psychiatric Severity Rating	-.11	.10	.27	.18	.18	.37	.62
Interview Psychiatric Severity Rating	-.22	.15	.14	.15	.05	.40	.50

N for ASI–MV = 109; N for ASI Interview = 50

[1]Short Form-12 Physical Components Scale (Ware et al., 1995), Note: Higher scores on this measure indicate less severity; [2]Social Adjustment Scale (SAS-SR; Weissman & Bothwell, 1976); [3]Michigan Alcoholism Screening Test (MAST; Selzer, 1975); [4]Drug Abuse Screening Test (DAST; Skinner, 1982); [5]Antisocial Behavior Checklist (ABC; Zucker et al., 1994).