Recent Status Scores for Version 6 of the Addiction Severity Index (ASI-6)

John S. Cacciola[*,#], **Arthur I Alterman**[*,#], **Brian Habing**[+], and **A. Thomas McLellan**[#]

[*]Treatment Research Institute, Philadelphia, PA, USA

[#]University of Pennsylvania School of Medicine, Philadelphia, PA, USA

[+]University of South Carolina, Department of Statistics, Columbia, SC, USA

Abstract

Aims—To describe the derivation of Recent Status Scores (RSSs) for Version 6 of the Addiction Severity Index (ASI-6).

Design—118 ASI-6 recent status items were subjected to nonparametric item response theory (NIRT) analyses followed by confirmatory factor analysis (CFA). Generalizability and concurrent validity of the derived scores were determined.

Setting and Participants—607 recent admissions to variety of substance abuse treatment programs constituted the derivation sample; a subset (N = 254) comprised the validity sample.

Measurements—The ASI-6 interview and a validity battery of primarily self-report questionnaires that included at least one measure corresponding to each of the seven ASI domains were administered.

Findings—Nine summary scales describing recent status that achieved or approached both high scalability and reliability were derived; one scale for each of six areas (medical, employment/ finances, alcohol, drug, legal, psychiatric), and three scales for the family/social area. Intercorrelations among the RSSs also supported the multidimensionality of the ASI-6. Concurrent validity analyses yielded strong evidence supporting the validity of the six of the RSSs (Medical, Alcohol, Drug, Employment, Family/Social Problems, Psychiatric). Evidence was weaker for the Legal, Family/Social Support and Child Problems RSSs. Generalizability analyses of the scales to males versus females and whites versus blacks supported the comparability of the findings with slight exceptions.

Conclusions—The psychometric analyses to derive Addiction Severity Index-6 Recent Status Scores (RSSs) support the multidimensionality of the ASI-6 (i.e., the relative independence of different life functioning areas), consistent with research on earlier editions of the instrument. In general, the ASI-6 scales demonstrate acceptable scalability, reliability and concurrent validity. While questions remain about the generalizability of some scales to population subgroups, the overall findings coupled with updated and more extensive content in the ASI-6 support its use in clinical practice and research.

Correspondence and requests for the ASI-6 and administration and scoring materials as well as training opportunities to: John S. Cacciola, Treatment Research Institute, 600 Public Ledger Building, 150 S. Independence Mall West, Philadelphia, PA 19106-3414, USA. jcacciola@tresearch.org.

Keywords

Addiction Severity Index; ASI; reliability; validity; measurement

Introduction

This paper describes the derivation of Recent Status Scores (RSSs) for Version 6 of the Addiction Severity Index (ASI-6). The ASI was developed in the late 1970's and the ASI-3 was the first widely distributed version [1]. A slightly modified version, the ASI-5, was introduced in 1992 [2], but is essentially the same instrument that was developed over 30 years ago. It was developed within the U.S. Veterans Administration (VA) substance abuse treatment system for primarily male alcohol and opioid dependent patients, the bulk of that population at the time. Nevertheless, the ASI is probably the most widely used clinical and research assessment instrument for individuals with substance use disorders (SUDs). It is used internationally to assess individuals with all types of SUDs in numerous settings and studies. The extensive use of the ASI in populations for which it was not originally tested, changes in the field and accumulation of knowledge that have taken place over the years, and research that has revealed limitations to the instrument led us to undertake a major revision [3, 4].

This paper will not detail the changes to the ASI that resulted in the ASI-6, as these are described elsewhere [4], but an overview of the new instrument and necessary information to contextualize the current research are presented. It is important to emphasize that our intention was to retain important hallmarks and strengths of the ASI while remedying gaps in content, and psychometric deficiencies [5]. Accordingly, three hallmarks of the ASI have been retained. We consider the ASI's multidimensional assessment to be a defining feature. Thus, the seven problem areas (i.e., medical. employment/support, drug, alcohol, family/social, legal and psychiatric) are generally preserved. Nonetheless, the content has been updated in order to fill recognized gaps, reflect current knowledge about SUDs and their treatment, and align items with current national databases and standards. Second, the ASI-6 remains comprised of mostly so-called "objective" (i.e., verifiable) items that document type, duration, and frequency of problems. These items are still supplemented with two "subjective" [i.e., 5-point (0-4), Likert–type, patient-rating scale] items that query the respondent regarding 'how troubled/bothered' by and 'how important treatment' is for problems in most areas. The lifetime and past-30 day time frames remain the primary assessment intervals. However, a criticism has been that the 30-day time frame is: too brief to adequately assess baseline functioning, and to support cost analyses; and too imprecise to assess recent problem acuity. To address these limitations, the ASI-6 includes a 6-month time frame and a date of last occurrence probe for key items.

Two general modifications were made to facilitate the interview process itself. First, although the ASI was originally designed as a semi-structured interview, the ASI-6 provides more structured wording of items to simplify interviewer training and increase reliability. Second, since the revised instrument is more comprehensive than its predecessors, in order to collect more information but keep the interview to less than an hour, the ASI-6 uses screening questions with "skip-outs" – a strategy generally not used in earlier ASIs.

The original ASI summary scores for recent severity are the Composite Scores (CSs) which were based on both rational and empirical methods. Given the considerable evidence for their validity, the CSs represented a formidable accomplishment for their time [3, 6-8]. However, they were not standardized, resulting in skewed scores and different distributions

across problem areas. Also, in some areas the internal consistency/reliability of the CSs were not entirely adequate [9, 10].

In an effort to derive a set of more psychometrically sound measures of recent status, exploratory factor analysis (EFA) was applied to the ASI-5's recent items [11]. Five of the seven ASI areas yielded summary scores (Evaluation Indices) that were stable and generalizable across subgroups, and maintained using confirmatory analyses. The medical and employment areas failed to meet these criteria. Due to this limitation, and methodological and theoretical considerations concerning EFA, we performed analysis of the ASI-5 recent items using nonparametric item response theory (NIRT) analyses on data from 2142 cases in studies conducted at our Center [12]. These yielded robust summary indices in each of the seven ASI areas. The present analysis of the ASI-6 employs NIRT in combination with independent confirmatory non-linear factor analysis (CFA) to derive summary measures of recent functioning.

Methods

Participants (Table 1)

A total of 607 patients completed the ASI-6 within one week of entry into one of 10 substance abuse treatment programs (7 community & 3 VA; alternatively, 4 outpatient, 4 inpatient & 2 methadone maintenance). A subset of 252 participants from 6 programs (3 community & 3 VA; alternatively, 2 outpatient, 2 inpatient, & 2 methadone maintenance) were administered a battery of instruments for validity analyses. All participants gave informed consent to participate in this IRB approved study.

Assessments

Participants were administered the ASI-6 and validity battery (described below) by a trained research technician in one session. The battery was designed to include at least one measure corresponding to each of the seven ASI domains. Self-report questionnaires were selected for efficiency, and to attenuate potential biases of a single technician administering both the ASI-6 and validity battery as interviews. A technician administered the questionnaires one at a time to participants by reviewing the instructions and reporting period, and answering questions. A final validity measure was derived from arrest data from the Pennsylvania Department of Corrections.

The ASI-6

An overview of the ASI-6 was presented above, information regarding recent status items in the instrument's seven nominal domains is now provided. The focus is on items with responses that are both directional (i.e., have a negative or positive valence and are not merely descriptive) and can change over time (because the resultant scores are intended to not only measure status at a single point but also to measure change with administrations of baseline and follow-up ASIs). These were the items considered for psychometric analyses.

Medical

There are 11 items in the ASI-6 that address current or past 30-day medical status. Generally, these recent medical items address physical health, medical problems, and related treatment service utilization.

Employment/Finances

Recent employment and financial support is assessed with 23 ASI-6 items. A total of 19 items were included for psychometric analyses as in two instances individual items were

combined (i.e., to create 2 'combination items') due to low base rates at the individual item level and similarity of constructs measured by the items. Recent items obtain information on employment status, barriers to employment and indicators of employability, employment income, other types of income, other indicators of personal finances (e.g., debt in arrears), and homelessness.

Drugs and Alcohol

The ASI-6 includes 45 items related to recent alcohol and other drug use, problems, and service utilization (including 2 items on tobacco use and gambling). There were 2 combination items (sum of all days of drug use; total number of drugs abused), thus 30 items were included for analysis.

Legal

Recent illegal activities and involvement in the criminal justice system are assessed with 16 items. A total of 11 items were included for psychometric analyses as 2 combination items were created.

Family/Social

The recent items that address family/social status on the ASI-6 were included to cover: adult relationships, which include problems and support (30 items); use of free time (4 items); problems and needs regarding minor children (8 items); and the occurrence of and response to trauma/victimization (7 items, which arguably straddle both the family/social and psychiatric domains). Of these 49 items, the psychometric analyses included 30, because there were 7 combination items.

Psychiatric

There are 21 items that assess the presence or absence, and independence or substance-relatedness of a variety of recent specific psychiatric symptoms, as well as associated distress, impairment, and service utilization. A total of 17 items were included for psychometric analyses as 1 combination item was created.

To summarize, 118 items that were either individual ASI-6 items or derived combination items were subjected to psychometric analysis to derive recent summary scales. This set of 118 items included the content of the total 165 recent status items that were both directional and subject to change over time.

Item Recoding

In addition to creating some combination items, it was necessary to recode continuous level data obtained for many items (e.g., frequencies of behaviors/experiences in the past 30 days, dollars, days ago an event occurred) prior to submission to NIRT analysis. The most common recode was a 5-point ordinal, Likert-type format (0 = 0; 1 = 1-5; 2 = 6-15; 3 = 16-25; 4   26) for past 30 day behaviors. Since actual frequencies and ranges of the responses varied among items, different categories were sometimes used to provide more uniform response frequencies among the categories. Nonetheless, the number of categories never exceeded five and occasionally dichotomous recodes were most appropriate. Although there was some arbitrariness to our recoding approach, our choices were informed by the actual distribution of each item's responses as well by the rational and clinical meaning of the categories. For example, the most common recoding choice allowed for an easy interpretation of 'not at all'; 'several times, or about once a week'; … to 'almost every day'. Finally, responses were reverse coded as necessary to maintain consistency within and across potential scales.

External Validity Measures

A validity battery was assembled to yield at least one measure for each ASI area.

Short-Form Health Survey (SF-12)

Recent medical and psychiatric problems were assessed using the SF-12, Version 2 [13] derived from the larger SF-36 questionnaire [14, 15]. The SF-36 and SF-12 measures daily life problems related to physical and mental difficulties and yields normalized standardized ($M = 50$, $SD = 10$) component summary scores for each [13, 14, 15]. Calsyn and colleagues found correlations of -0.61 between the SF-36 physical score and the ASI-5's medical CS and -0.73 between the mental score and the psychiatric CS [16]. Similarly, the correlations between the SF-36's physical and mental scores and the ASI-5's NIRT-derived recent medical and psychiatric summary scores were -0.56 and -0.52, respectively [12]. Studies have shown that the SF-12 physical and mental scales reproduce 90% of the variance in the SF-36's corresponding measures [13, 17]. Effect size differences between the scores of SF-36 and SF-12 are small [18]. The 4-week version of the SF-12 was used. Higher scores indicate better functioning.

SCL-10R

This 10-item measure of psychopathology [19] was derived from an analysis of the more consistent item loaders from a number of factor analytic studies of the SCL-90 [20]. The internal consistency of the SCL-10R was 0.87 in a large VA sample of PTSD patients. Its correlation with the General Severity Index of the SCL-90 was 0.95 and it had moderate to high correlations (0.55) with other measures of psychopathology (i.e., depression, anger, anxiety, PTSD). Additionally, the correlation of the SCL-6 (a more abbreviated version) with the ASI-5 psychiatric CS was 0.69. The SCL-10R assesses a 1-week timeframe. Higher scores indicate greater psychopathology.

Social Adjustment Scale Self-Report [SAS-SR; 21]

This is a longstanding measure of family, social and employment functioning that has been shown to have good inter-rater reliability (ICC = 0.62) [21]. Test-retest reliability (ICC = 0.80), internal consistency (α = 0.74) and concurrent validity have been demonstrated [22]. In the later case, overall social adjustment summary scores in a normal and three patient samples (depressed, schizophrenic, alcoholic) were highly correlated with measures of depression (CES-D; $r's$ = 0.49 - 0.84) and general psychopathology (SCL-90; $r's$ = 0.59 - 0.84) [22]. Additionally, significant differences in scores among the four samples support the instrument's discriminant validity [22]. The SAS-SR has been used use as a measure of concurrent validity for the ASI-3 [6, 23]. The timeframe is the past two weeks with higher scores indicating poorer functioning.

Employment adjustment in the SAS-SR is defined as work functioning outside the home, at home, or as a student and assessed with 18 of the instruments 54-items. Since these situations applied to few participants, we used a single SAS-SR item which indicated whether or not the respondent worked for pay in the past two weeks.

Family/social functioning within the SAS-SR is assessed with 35 items in 5 areas: social and leisure activities, extended family, marital role, parent, and immediate family unit. Most items are rated on a 5-point scale ranging from 1, the most positive response, to 5, the most negative. Therefore high scores indicate greater problems. Typically the scores for the items in each of the area are separately summed and divided by the number of items within the area to yield a score. Since the majority of participants were not married nor parents, few participants responded to these sections. To derive a global measure of family/social functioning, when the respondent completed 3 or more of the 5 areas we summed the mean

of the area scores and divided that by the number of areas. This resulted in a summary score for 217 participants [out of 252 (86.1%)].

The Short Index of Problems [SIP; 24]

This 15-item measure of consequences of alcohol use is derived from the 45-item Drinker Inventory of Consequences [DrInC; 24]. Miller and colleague constructed the SIP with three items from each of the DrInC's five subscales and found high internal consistency (α = . 0.81) and 2-day, test retest reliability (r = 0.94) for the lifetime SIP total score. A later study using a 3-month SIP found internal consistency of 0.79, 3-month, test-retest stability of 0.74, and a correlation of 0.87 between the total score and the 30 remaining Drinc items [25]. Our own 3-month SIP data analysis in this sample yielded one factor (α= 0.98) [26]. Higher scores indicate more adverse consequences.

The Short Index of Problems - Drugs [26]

Tonigan and Miller, in response to a paucity of brief measures to assess the consequences of recent drug use constructed a measure for both alcohol and drug use by essentially adding the term 'or drug use' after 'drinking' in each DrInC item [27]. This instrument, the Inventory of Drug Consequences [InDUC; 27], and several briefer variants have good psychometric properties [28, 29]. We constructed a drug version of the SIP (SIP-D) to assess the effects of drug use only for the current study [26]. Factor analysis revealed one primary factor with high internal consistency (α = 0.97). A 3-month time frame was used.

California Psychological Inventory-Socialization Scale [CPI-So; 30]

CPI-So is a measure of socialization, social judgment, and normative behavior which yields a summary measure of antisociality. This self-report questionnaire contains 46 items that address adolescent and adult experiences, behaviors and beliefs. The psychometric properties and validity of the CPI-So have been excellent in several populations [30, 31]. A study of convergent validity suggests that this scale is a good measure of antisociality in substance abusers [32]. Predictive validity has also been demonstrated in substance abuse patients [33, 34]. A lower score is indicative of greater severity, reflecting poorer social judgment, less empathy and less conformity with social norms.

Criminal Records

State criminal records were used to derive a variable to reflect the sum of documented arrests in the two years prior to study intake. This variable was employed in our earlier NIRT analysis of the ASI-5 [12].

Psychometric Analyses

Like most standard statistical methodologies, factor analysis (EFA and CFA) has difficulties with data that contain a mixture of binary, Likert scale, and continuous responses, a condition that characterizes the ASI. In particular, linear factor analysis of binary or Likert scale items can lead to both incorrect dimensional structures being recovered and hypothesis tests with inaccurate p-values and inflated type I error rates. This is true even when tetrachoric and polychoric correlations are used [e.g., 35, 36, 37]. Alternatively, item response theory (IRT) methods are designed to model the relationship between dichotomous (e.g., binary) and polytomous (e.g., Likert scale) responses to a questionnaire and the underlying trait(s). Its foundational work in the 1960s [38, 39], IRT is currently used to analyze virtually every large-scale standardized educational test in the U.S. and has had growing prominence in analyzing health instruments. IRT models can also be re-parameterized to take the form of a nonlinear factor analytic model [e.g., 40, 41]. The non-linearity allows for the correct recovery of a unidimensional scale without the restriction of

normality [e.g., 42]. Unfortunately, even in the dichotomous case the standard (parametric) multidimensional IRT implementations have difficulty in recovering correct high dimensional structures when fit in an exploratory fashion [e.g. 43, 44].

As an alternative, the nonparametric IRT based conditional covariance [45] and Mokken scaling methods [46] are able to recover the underlying dimensional structure of a multidimensional instrument without needing to fit the underlying model. Both methods were applied in tandem to the ASI-5 [12] by analyzing a pre-determined group of items within each ASI area separately. This approach did not appear to be justified for the ASI-6 where major changes in area and item content strongly suggested a more conservative approach. Thus, our analysis is on the entire set of recent items rather than on items in one designated area at a time. This large number of extremely varied items made the standard (graphical based) output of the conditional covariance based methods unwieldy and the underlying multidimensional structure required by the methods [e.g., 47] seem untenable. Mokken scaling does not suffer from either limitation and was implemented.

Mokken scaling [46] is a nonparametric IRT (NIRT) based method for constructing scales of dichotomous or polytomous items that measure unidimensional latent traits. It can be used in an exploratory fashion to refine a large set of items into dimensionally distinct scales or in a confirmatory fashion to verify the dimensional homogeneity of a particular predetermined set of items [48]. The basic foundational idea of Mokken scaling is that a respondent's total score on a scale is only easily interpretable if the scale follows a monotone homogeneity IRT model (MHM) with the properties of unidimensionality, local independence, and monotonicity. In particular these are sufficient so that, on average, a higher observed score corresponds to a higher value on the latent trait. This condition is tied to the observed data by the fact that a scale satisfying the MHM will have (generally large) positive covariances between all pairs of items.

Mokken scaling is thus based on using appropriately rescaled covariances, called scalability coefficients. The pairwise item scalability coefficients, abbreviated H_{ij} for items i and j, are used to verify that each of the items in the potential scale are related to each of the other individual items in that scale. This occurs if each H_{ij} value is greater than 0. The item scalability coefficient, H_i, measures the relationship between the particular item and the rest of the items on the proposed scale as a whole. This value is typically desired to be above a minimum value of 0.3 to indicate that each individual item adequately measures the construct measured by the whole scale, and is summarized as the minimum Hi (minHi). Finally, the overall scalability coefficient, H, measures the homogeneity of the scale. An H of 0.3 to 0.4 is taken to indicate a weak scale, 0.4 to 0.5 a moderate scale, and 0.5 to 1.0 a strong scale [48].

The information provided by the scalability coefficient H is complementary to that provided by a measure of reliability such as Cronbach's alpha coefficient (α). Alpha provides a measure of the replicability of a score on a scale (i.e., an estimate of the correlation of the score on the instrument with the score on a parallel instrument). It will be large for a long instrument, even if that instrument measures several distinct constructs. The overall scalability coefficient (i.e., H) indicates that a scale measures a single construct, without providing insight into how precise that measure will be. A single item pair can form a scale with an H of 0.5, for example. A quality scale will thus generally have both a high scalability (i.e., all H_{ij} 0, H_i 0.3, and H 0.5) and high reliability (α 0.70). As all quantities are measured with some error and items were included for substantive reasons, some final scales may not fully meet these targets.

Sijtsma and Molenaar [48] describe a sequential method for refining a large multi-faceted instrument into separate unidimensional scales. The method begins by choosing the item pair with the largest H_{ij} to form the base of the proposed scale. Items are then added one at a time, if they have a positive pairwise H_{ij} with each of the earlier items, and a large enough item scalability, H_i, with this previously chosen set of items as a whole (typically 0.3). This process continues until none of the remaining candidate items meet the criteria. Once the first scale is formed, the process is repeated using all of the items not already selected. Our implementation of exploratory Mokken scaling differed from Sijtsma and Molenaar's approach in that we allowed for reverse coded items as indicated by the H_{ij}, and conducted a more thorough search for possible scales. This more thorough search was done by considering every pair of items as the basis for a possible scale instead of simply those with the highest H_{ij} coefficient. In many cases this incremental procedure produced several similar versions of a scale, differing only by inclusion or exclusion of a few items. The decision between these possible scales was informed by the insights of content area experts (the ASI-6 Core Development Group led by Alterman, Cacciola and McLellan, see Cacciola et al., submitted paper). As noted, this exploratory process was performed for 118 items that were primarily individual ASI-6 items but included derived combination items as well. This item set represented the content of 165 ASI-6 recent status items.

The scalability coefficients described above were supplemented by α. A value of 0.70 is often recommended as a threshold for acceptability for this statistic [49]. A high-quality subscale should thus have large values of minH_i, H, and α indicating that it reliably measures a homogeneous underlying latent factor.

Validity Analyses

Two forms of validity analyses were undertaken. One analysis examined the intercorrelations between Recent Status Scores (RSSs) in the different problems areas to determine the extent of independence of the areas. Second, the correlations between these indices and nine external measures of corresponding functioning were examined. Prior to the intercorrelational analyses of the RSSs, weighted total scores of the items in each scale were transformed into area conversion *T*-scores [8, 50]. We then compared intercorrelations between the ASI-6 RSSs with the expectation of generally low correlations among the areas. Second, the RSSs in each of the areas were compared with external validity measures, with the expectation that the RSS would be most highly related to its corresponding external validity measure(s), as compared to external measures for other problem areas. For the external validity analyses, bivariate correlations were performed between the standardized scores for the RSSs and the nine external validity measures.

Results

Overall 9 summary scales to assess recent status were derived using Mokken scaling methods that achieved or approached both high scalability and high reliability. One scale was derived for each of six areas, medical, employment/finances, alcohol, drug, legal, and psychiatric, and three scales were derived for the family/social area. The constituent items that comprise each scale are shown in Table 2 and serve to describe their content.

A 7-item recent summary Medical scale emerged ($H = 0.56$; min$H_i = 0.36$; $\alpha = 0.82$) that included a medical service utilization item, the respondent's ratings of physical health, medical problem severity, pain, and treatment need, as well as two problem frequency items.

Regarding employment and finances, an Employment scale was identified ($H = 0.87$, min$H_i = 0.70$, $\alpha = 0.89$) that contained four items all dealing directly with employment. Of the

many items related to financial viability such as income from other sources, debt, etc., or to employability, none ended up in the Employment scale or resulted in a separate scale(s).

Scales were derived for both Alcohol (11 items: $H = 0.71$; $\min H_i = 0.60$, $\alpha = 0.93$) and Drug (12 items: $H = .0.73$, $\min H_i = 0.63$, $\alpha = 0.94$) items. The scales were composed of essentially the same items except for their reference to alcohol versus drugs, respectively. These items included frequency of use and last use, specific types of problems (e.g., craving, withdrawal) and overall frequency of problems, money spent on the substance(s), and patient ratings of 'how troubled/bothered by' and 'how important is treatment for' substance use.

A 6-item Legal scale emerged ($H = 0.54$, $\min H_i = 0.28$, $\alpha = 0.78$). Of the six items, five queried frequency of various illegal activities and one amount of money made from illegal activities. No items regarding current involvement in the criminal justice system met criteria for inclusion.

The analysis produced three Family/Social scales; Family/Social Problems, Family/Social Support, and Child Problems. A 5-item Family/Social Problems scale ($H = 0.49$, $\min H_i = 0.46$, $\alpha = 0.71$) focuses on a variety of problems in adult relationships (i.e., with spouse/ partner, other adult relatives, close friends). Problems include trouble getting along, arguments, and physical conflict with these adults as well as patient ratings of 'how troubled/bothered by' and 'how important is treatment for' these problems. A 4-item Family/Social Support scale was also derived ($H = 0.54$, $\min H_i = 0.40$, $\alpha = 0.78$) that includes various types of support (e.g., spend time with, can count on for help) that the respondent received from adult relationships. A 5-item Child Problems scale ($H = 0.51$, $\min H_i = 0.35$, $\alpha = 0.73$) describes the number of children living with the respondent who have a serious problem and their need for services for these problems, trouble the respondent has getting along with children living with him/her, and patient ratings of the importance of counseling for such problems and need for childcare services.

The Psychiatric scale has 17 items ($H = 0.48$, $\min H_i = 0.34$, $\alpha = 0.87$). These include the presence or absence and substance-relatedness of numerous specific psychiatric symptoms and the most recent occurrence of the more severe symptoms (e.g., suicidality), overall frequency of symptoms, frequency of related impairment and days hospitalized, as well as patient ratings of associated distress and treatment need. Patient ratings of trauma-related distress and treatment need are also included.

Nonlinear Confirmatory Factor Analysis (CFA)

As noted above, parametric multidimensional IRT models for polytomous items (equivalent to a nonlinear factor analytic model) have a variety of difficulties when used in an exploratory fashion. When used in a confirmatory fashion, however, they may provide some additional evidence that the various scales chosen are well defined. CFA using the nine scales identified above was performed in M-Plus [51] using the categorical option and the weighted least squares estimator (WLSMV). Examining the fit of the CFA gave a CFI of 0.930, TLI (Tucker-Lewis) of 0.956, and RMSEA of 0.083. The TLI is within the recommended range found by Yu [52] for dichotomous item CFAs, while the CFI and RMSEA are slightly outside the desired range. Given the potential lack of fit due to the graded response model being the incorrect parametric model, this does not seem to indicate a serious lack of fit. All of the factor loadings were significant and ranged between 0.39 and 0.99, with only 3 of the 71 loadings below 0.56. Most of the correlations between latent constructs were low with 75% below 0.27 and none above 0.58. This indicates that the items were correctly assigned to the scales, and that the scales measured distinct constructs.

Generalizability to Sociodemographic Subgroups

A key question in the use of statistical models is the stability of their solutions for different subpopulations. Two studies of this stability were conducted. The first was a comparison of the item-pair H_{ij} statistics between demographic groups (i.e., male vs. female, black vs. white) to provide some assurance that similar potential scales would have been selected, if the exploratory analysis had been performed separately on the different subgroups. The H_{ij} for each item pair was calculated separately for the contrasting groups, and classified as being less than 0.3, between 0.3 and 0.4, between 0.4 and 0.5, and greater than 0.5. The percentage of item pairs where the classifications were within one level of each other were counted, as was the percentage of item pairs where one classification was >0.5 and the other was <0.3. Permutations tests (one-sided) were then conducted to see if these percentages were statistically different than would be expected by chance. For males versus females, the p-values of 0.211 and 0.358 gave no significant evidence that the differences in H_{ij} classification were greater than would be found by chance. In the case of respondents identifying themselves racially as black (and no other race) and white (and no other race), the p-values were <0.001 and 0.002 indicating that the pair-wise scalability coefficients performed differently between the two racial groups. In examining the statistics used to calculate the p-values, the actual percentages of being classified within one level of each other, or of being classified >0.5 for one and <0.3 for the other, for the blacks versus whites were not far from the range simulated in the permutation tests. Specifically, 91.57% of the actual pairs were within one classification versus a simulated range of 91.66% to 93.37%, and 4.99% of pairs showed a >0.5 versus <0.3 discrepancy as compared to a range of 3.60% to 5.01% in the re-sampled values. Thus, while the results were statistically significant for blacks versus whites, the actual magnitude of the difference between them was small, indicating that scales developed separately on those two populations would not differ too much beyond what would be expected by chance.

The second study examined whether the final scales would have acceptable scalability and reliability if evaluated on each of the four demographic groups separately (i.e., male, female, black, white). The Medical, Employment, Alcohol, Drug, and Family Support scales met all of the targets (H 0.5, minH_i 0.3, minH_{ij} 0, α 0.7) for all four subgroups. In most cases, the Legal, Family Problem, Problems with Children, and Psychiatric scales missed the targets in a similar manner to the scales found using the entire sample (see Table 3) and any discrepancies were well within what might be expected from the smaller sample sizes of the subgroups (see below). The scalability and reliability properties for the nine scales thus seem fairly stable across demographic subpopulations.

Sensitivity to Sample Size

Our earlier work on the use of Mokken scaling on the ASI-5 [12] found that the scalability coefficients tended to be fairly stable for sample sizes 500. An estimated 96.5% of item pairs with a scalability coefficient of less than 0.3 would be successfully classified as such, and 87.8% of those with a scalability coefficient of at least 0.5 would be successfully classified. The re-sampling methodology used for that particular result required the larger sample size available in that study (n = 2,142) and was thus not replicable on this ASI-6 data set. However, the similarity of the actual items between the ASI-5 and ASI-6, and of the distribution of item-pair H_{ij} classifications (<0.3, 0.3 to 0.4, 0.4 to 0.5, > 0.5) between the two samples provides some reassurance that the Mokken scaling procedure performed well for this ASI-6 sample of 607.

Derivation of Standardized *T*-scores for the Recent Status Scores (RSSs)

As noted, RSSs were standardized prior to submission to analysis. First, the individual items constituting a RSS were weighted before summing in order to counteract the fact that items

with more response categories will generally have a much larger influence on total scores than those with fewer. The binary item scores were multiplied by the square root of 8, those scored 0-2 by the square root of 3, and those scored 0-3 by the square root of 1.6. These are the multiples required to give all of the items the same variance (and thus the same weight) as a 0-4 item in the case where all of the possible responses are equally endorsed.

An alternative to this weighting scheme would be to use factor scores from the non-linear CFA. These scores would typically be found using a Bayesian implementation [e.g. 42] requiring more complex programming or customized software, and their exact values would depend on the original sample of subjects used [e.g. 53]. Sum scores are typically very highly correlated to this type of factor score [e.g. 54] and the calculated values are not affected by small changes in sample characteristics.

The weighted total scores for each RSS were then transformed into area-conversion T-scores [8, 50]. Unlike linear T-scores, area conversion maintains interpretive equivalence across scales by assigning standard scores according to corresponding raw-score percentiles as expected under the normal curve. The T- score for a given observation is the quantile of a normal distribution with $M = 50$ and $SD = 10$ that corresponds to the same percentile as the observation (e.g., the observation at the 2.5th percentile is replaced with 30.4, and the median is replaced with 50). Finally, in all areas higher scores indicate greater problem severity.

Validity Findings

Intercorrelations between the ASI-6 RSSs—The intercorrelations among the nine RSSs were examined using bivariate correlations. Ideally, low correlations are expected between the scores in the different problem areas. The correlations were generally low, supporting the independence of the derived scales (Table 4). However, psychiatric and drug problems were each associated with problems in other areas. The correlations of psychiatric problems with medical, family/social problems, drug, and alcohol scores were 0.44, 0.30, 0.33 and 0.42, respectively. The correlations of drug problems with legal and family/social problems were 0.40 and 0.32 respectively, and as noted, 0.33 with psychiatric problems. The correlation of alcohol problems with other ASI-6 areas was lower than that of drug problems with other areas, with only the correlation between alcohol and psychiatric problems reaching 0.30.

Concurrent Validity—The validity findings for the bivariate correlations between the ASI-6 RSSs and corresponding and noncorresponding external validity measures are shown in Table 5. These findings generally supported the concurrent validity of the ASI-6 summary scores.

Medical—The Medical RSS was correlated -0.64 with the SF-12 physical component score. It was correlated to a moderate, but lesser, degree with the SF-12 mental score (-0.34) and the SCL-10R (0.44) measure of psychopathology. The two latter relationships parallel that found between the ASI-6 Medical and Psychiatric scores, described above.

Psychiatric—The ASI-6 Psychiatric score was correlated -0.61 with the SF-12 mental component score and 0.68 with the SCL-10R measure. Moderate, but lower, relationships were also found between the Psychiatric RSS and SIP-D measure of drug consequences (0.37) and CPI-So measure of antisociality (-0.34).

Alcohol—Alcohol problems as measured by the Alcohol RSS were correlated 0.68 with the SIP measure of alcohol consequences. There were no other remarkable relationships with

other external validity measures. A correlation of 0.29 with the SCL-10R was the next highest correlation. Interestingly, the correlation of ASI-6 alcohol problems with the SIP-D was 0.01, not dissimilar from the .03 correlation between the ASI-6 Alcohol and Drug RSSs.

Drug—The ASI-6 Drug score was correlated 0.61 with the corresponding validity measure, the SIP-D. Its correlation with the noncorresponding external validity measures were unremarkable, the highest being with measures of psychiatric problems (SCL-10R, 0.30; SF-12 Mental, -0.27).

Employment—The Employment RSS was correlated 0.76 with the one-item validity measure drawn from the SAS-SR which reflected whether or not the respondent recently worked for pay. None of the correlations with any other validity measures reached 0.15.

Legal—The ASI-6 Legal score was not highly correlated with the CPI-So (-0.14) nor was it highly correlated with the number of arrests in the prior two years (0.15). Nonetheless, these were the strongest correlations of the Legal RSS to any external validity measures with the exception of the ASI-6 Legal score to the SIP-D (0.28).

Family/Social—A correlation of 0.40 was found between the ASI-6 Family/Social Problems RSS and the SAS-SR family/social functioning summary score. The correlation between the SAS-SR summary score and the ASI-6 Psychiatric RSS was, however, slightly higher (0.44). The ASI-6 Family/Social Problems score was moderately correlated with the external psychiatric validity measures (0.36). A correlation of -0.34 was obtained between the Family/Social Support RSS and the SAS-SR summary score indicating more support was associated with more problems. No other remarkable correlations were obtained between the ASI-6 Family/Social Support score and any other validity measures (all   0.15). The Child Problems RSS was essentially uncorrelated (-0.04) with the SAS-SR summary score. None of the correlations with other validity measures were noteworthy (all   0.16).

Discussion

Our analyses supported the multidimensionality for Version 6 of the Addiction Severity Index (ASI-6) similar to that which had been found in earlier versions of the instrument. The psychometric analyses to derive Recent Status Scores (RSSs) for the ASI-6 yielded nine primary scales, one for each of six areas (i.e., medical, employment/finances, alcohol, drug, legal, psychiatric) and three in the family/social area. Six scales, Medical, Employment, Alcohol, Drug, Family/Social Support and Child Problems, met all initial conditions for a strong quality scale (all H_{ij} 0, H 0.5, miniH_i 0.3, α 0.70). The remaining three scales approached these values but fell a bit short on one metric (i.e., Legal, miniH_i = .28; Family/Social Problems H = 0.49; Psychiatric, H = 0.48) and can conservatively be considered moderate quality scales. As all quantities are measured with some error, and occasionally items were included based on substantive reasons to more fully represent the construct under consideration, it is not surprising that some final scales slightly missed the desired thresholds. The CFA provided additional support for the scales indicating that the items were correctly assigned to the scales, and that the scales measured distinct dimensions. The intercorrelations among the nine RSSs further supported the multidimensionality of the ASI-6, revealing relative independence of the different life functioning areas. The stronger relationships were those that also had been identified with previous versions of the ASI [6, 8, 12].

Analyses of the generalizability of the derived scales to males versus females and whites versus blacks generally supported the comparability of the findings with some slight exceptions. The comparison of the item-pair H_{ij} statistics between demographic groups

provided some assurance that similar scales would have been selected, if the exploratory analysis had been performed separately on the subgroups. For males versus females there was no significant evidence that the differences in H_{ij} classification were greater than would be found by chance. For blacks versus whites, while the results were statistically significant the magnitude of the difference between them was small. The examination of whether the final scales would have acceptable scalability and reliability if evaluated on each of the four subgroups separately revealed that in most cases, the Legal, Family/Social Problems, Problems with Children, and Psychiatric scales missed the targets in a similar manner when using the entire sample and any discrepancies were well within what might be expected from the smaller sample sizes of the subgroups. Thus, questions do remain regarding generalizability to both the specific subgroups that were subjected to analyses and to other important subgroups. Future analyses on larger samples can provide new and more definitive data concerning the generalizability of the RSSs.

The concurrent validity analyses with the external measures yielded strong evidence supporting the validity of six of the RSSs (Medical, Alcohol, Drug, Employment Family/ Social Problems, Psychiatric). The external validity measures for these areas were arguably better than for those in the remaining areas. They assess recent status, are directly related to the corresponding ASI-6 area, and are psychometrically strong scales comprised of multiple items (with the exception of the single SAS-SR employment item). The evidence for concurrent validity was minimal in the Legal area. Possible reasons are that a history of documented prior arrests may be relatively insensitive regarding very recent illegal behavior as measured by the Legal RSS. Similarly, the CPI-So, a personality measure, may be insensitive to specific recent behaviors occurring over only a short period of time. It is encouraging that additional predictive analyses yielded a correlation of 0.34 between the Legal RSS and documented arrest in the two-year period following study intake. Regarding the Family/Social Support and Child Problems RSSs there were admittedly no direct external validity measures. A negative correlation between the ASI-6 measure of support and the global SAS-SR measure of family/social functioning was nonetheless unexpected. One possible explanation is that the Family/Social Support RSS's content consists largely of items that have to do with interpersonal contact. Thus the opportunity for more problems may exist for those respondents with lower scores (i.e., more support) than for those who are more isolated. Future studies that compare this RSS with direct measures of support (e.g., the Social Provisions Scales) [SPS; 55] would be informative and are warranted. There was also no direct external validity measure for the Child Problems RSS. It was not related to the summary score provided by the SAS-SR, likely because this measure is far too general to be sensitive to specific problems assessed by a Child Problems Scale. The uniformly low correlations of the ASI-6 Child Problem RSS with the external measures selected seem to neither support nor challenge its validity. Again, future studies that compare this RSS with direct corresponding measures are warranted.

Seven of the nine RSSs closely parallel the scales that measure recent functioning derived from earlier versions of the ASI (ASI-3/5) (i.e., Medical, Employment, Alcohol, Drug, Legal, Family/Social Problems, Psychiatric), and some content in each ASI-6 scale overlaps with content in the corresponding recent scale(s) from earlier versions. For these seven areas, it is therefore not surprising that the scalability and reliability, intercorrelations among the RSSs, and the concurrent/external validity results for the ASI-6 are roughly comparable to the results obtained with similar analyses using the ASI-3/5. This being said, a natural question is, "Why use the ASI-6 instead?" Specifically, "What are the advantages of the ASI-6 regarding the assessment of recent functioning?" There are several facets to our response:

First, the set of 118 items analyzed to derive the nine RSSs included the content of the total 165 ASI-6 recent status items. This is in contrast to the content of approximately 80 recent status items used to derive summary scales to measure recent functioning with earlier versions of the ASI (ASI-3/5). Moreover, in nearly all of these seven ASI-6 scales the number of items included exceeded those in the corresponding ASI-3/5 scales. Thus, the ASI-6 offers more comprehensive content in its scales than do those derived with earlier ASIs. Examples of the enhanced content versus earlier ASIs are additional items that query last use and specific types of problems (e.g., craving, withdrawal) in the alcohol and drug areas of the ASI-6. Also, in the psychiatric section the inclusion of items that assess the substance-relatedness for each specific symptom endorsed, the most recent occurrence of the more severe symptoms, days hospitalized for mental health problems, and patient ratings of trauma-related distress and treatment needs are items not included in earlier ASIs. In a related manner, the ASI-5 and ASI-6 recent psychiatric measures performed similarly in identifying co-morbidity, but the ASI-6 performed better for PTSD [56].

Second, there are numerous ASI-6 recent status items that are not in the earlier ASIs and have clinical and/or research value (e.g., homelessness, pregnancy, tobacco use, gambling) but did not emerge in the RSSs.

Third, in addition to the seven ASI-6 scales that parallel those found in earlier ASIs, it was possible to delineate two more specific summary measures in the ASI-6's family/social area (i.e., Family/Social Support, and Child Problems) due to its greater differentiation and expansion. In this regard, additional analyses yielded six additional preliminary scales in the family/social area; problems and support with the respondent's spouse/partner, other family, and friends. To varying degrees, these scales approached or exceeded the targets for quality scales, but since the content overlapped completely with the Family/Social Support and Family/Social Problems RSSs we decided that the more global measures would be among the set we consider primary ASI-6 summary scales.

It should be acknowledged that our inability to derive a more comprehensive or several more specific RSSs in the employment/finances area was a disappointment given the ASI-6's expanded coverage. In this regard, one limitation of the current study was that the study sample was not entirely representative of the larger population of SUD patients in this country. The majority of participants were unemployed African Americans treated in largely publicly funded programs. In particular, only a small minority was employed. It is possible that a more comprehensive or additional RSS would have emerged for the Employment/Finances area, if a greater proportion of those assessed had been employed and less financially marginal. Nonetheless, we included patients from multiple treatment organizations and modalities, and the sample may not be dissimilar in important ways from an urban, particularly inner city, U.S. treatment-seeking population.

In addition to the non-representativeness of the study sample, another limitation of the current analysis was the relatively small sample size for such a major undertaking. A larger and more sociodemographically diverse sample would yield more definitive findings, although we would anticipate substantial convergence between the results of the current study and those of a larger study.

In conclusion, our analyses supported the multidimensionality of the ASI-6 that had been found in earlier versions of the instrument (i.e., the relative independence of the important different functional areas) with quality scales that assess recent functioning. These positive findings with the ASI-6 coupled with its updated and more extensive content, and similar administration time compared to earlier ASIs support the use of the ASI-6 in clinical practice and research. Additionally, the current study has provided the foundation for future

research in order to further confirm or extend our findings. Translations of the ASI-6 into other languages have been accomplished and such research is ongoing in other countries (e.g., Brazil, Spain) [57, 58]. Specific future directions we plan to pursue are the derivation of lifetime summary scores, and comparison studies of the ASI-6 and the ASI-5.

Acknowledgments

References

1. McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients: the Addiction Severity Index. J Nerv Ment Dis. 1980; 168:26–33. [PubMed: 7351540]

2. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. The fifth edition of the Addiction Severity Index: historical critique and normative data. J Subst Abuse Treat. 1992; 9:199–213. [PubMed: 1334156]

3. Mäkelä K. Studies of the reliability and validity of the Addiction Severity Index. Addiction. 2004; 99:398–410. [PubMed: 15049734]

4. McLellan AT, Cacciola J, Alterman AI, Rikoon SH, Carise D. The Addiction Severity Index at 25: origins, contributions and transitions. Am J Addict. 2006; 15:113–24. [PubMed: 16595348]

5. Butcher JN. Revising psychological tests: lessons learned from the revision of the MMPI. Psychol Assess. 2000; 12:263–71. [PubMed: 11021149]

6. McLellan AT, Luborsky L, Cacciola J, Griffith J, Evans F, Barr HL, et al. New data from the Addiction Severity Index: reliability and validity in three centers. J Nerv Ment Dis. 1985; 173:412–23. [PubMed: 4009158]

7. McLellan AT, Cacciola JS, Alterman AI. The ASI as a still developing instrument: response to Mäkelä. Addiction. 2004; 99:411–2.

8. McDermott PA, Alterman A, Brown L, Zaballero A, Snider E, McKay J. Construct refinement and confirmation for the intake Addiction Severity Index scales. Psychol Assess. 1996; 8:182–9.

9. Alterman AI, Brown LS, Zaballero A, McKay JR. Interviewer severity ratings and composite scores for the ASI: a further look. Drug Alcohol Depend. 1994; 34:201–9. [PubMed: 8033757]

10. Hodgins DC, El-Guebaly N. More data on the Addiction Severity Index Severity Index: reliability and validity with the mentally ill substance abuser. J Nerv Ment Dis. 1992; 180:197–201. [PubMed: 1588339]

11. Alterman AI, McDermott PA, Cook TG, Metzger D, Rutherford MJ, Cacciola JS, et al. New scales to assess change in the Addiction Severity Index for the opioid, cocaine, and alcohol dependent. Psychol Addict Behav. 1998; 12:233–46.

12. Alterman AI, Cacciola JS, Habing B, Lynch K. ASI recent and lifetime summary indices based on nonparametric IRT methods. Psychol Assess. 2007; 19:119–32. [PubMed: 17371127]

13. Ware JE, Kosinski M, Keller SD. A 12-item Short Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996; 34:220–33. [PubMed: 8628042]

14. Ware JE, Kosinksi M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the medical outcomes study. Med Care. 1995; 33:AS264–AS79. [PubMed: 7723455]

15. Ware, JE.; Snow, K.; Kosinski, M.; Gandek, B. SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: The Health Institute; 1994.

16. Calsyn DA, Saxon AJ, Bush KR, Howell DN, Baer JS, Sloan KL, et al. The Addiction Severity Index medical and psychiatric composite scores measure similar domains as the SF-36 in substance-dependent veterans: concurrent and discriminant validity. Drug Alcohol Depend. 2004; 76:165–71. [PubMed: 15488340]

17. Jenkinson C, Layte R. Development and testing of the UK SF-12 (short form health survey). J Health Serv Res Policy. 1997; 2:14–8. [PubMed: 10180648]

18. Kontondimopoulos N, Pappa E, Niakas D, Tountas Y. Validity of SF-12 summary scores in a Greek general population. Health Qual Life Outcomes. 2007; 5:55–64. [PubMed: 17900374]

19. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull. 1973; 9:13–28. [PubMed: 4682398]

20. Rosen CS, Drescher KD, Moos RH, Finney JW, Murphy RT, Gusman F. Six- and ten-item indexes of psychological distress based on the Symptom Checklist-90. Assessment. 2000; 7:103–11. [PubMed: 10868247]

21. Weissman MM, Bothwell S. Assessment of social adjustment by patient self-report. Arch Gen Psychiatry. 1976; 33:1111–5. [PubMed: 962494]

22. Weissman MM, Prusoff BA, Thompson WD, Harding PS, Myers JK. Social adjustment by self-report in a community sample and in psychiatric outpatients. J Nerv Ment Dis. 1978; 166:317–28. [PubMed: 650195]

23. Kosten TR, Rounsaville BJ, Kleber HD. Concurrent validity of the Addiction Severity Index. J Nerv Ment Dis. 1983; 171:606–10. [PubMed: 6619823]

24. Miller, WR.; ToMiller, WR.; Tonigan, JS.; Longabaugh, R. Project Match Monograph Series. Rockville, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism; 1995. The Drinker Inventory of Consequences (DrInC): an instrument for assessing adverse consequences of alcohol abuse.

25. Feinn R, Tennen H, Kranzler HR. Psychometric properties of the Short Index of Problems as a measure of recent alcohol-related problems. Alcsm Clin Exp Res. 2003; 27:1436–41.

26. Alterman AI, Cacciola JS, Ivey MA, Habing B, Lynch KG. Reliability and validity of the alcohol SIP and a newly constructed drug Short Index of Problems. J Stud Alcohol Drugs. 2009; 70:304–7. [PubMed: 19261243]

27. Tonigan JS, Miller WR. The Inventory of Drug Use Consequences (InDUC): test-retest stability and sensitivity to change. Psychol Addict Behav. 2002; 16:165–8. [PubMed: 12079257]

28. Bender RE, Griffin ML, Gallop RJ, Weiss RD. Assessing negative consequences in patients with substance use and bipolar disorders: psychometric properties of the Short Inventory of Problems (SIP). Amer J Addict. 2007; 16:503–9. [PubMed: 18058418]

29. Blanchard KA, Morgenstern J, Morgan TJ, Labouvie EW, Bux DA. Assessing consequences of substance use: psychometric properties of the Inventory of Drug Use Consequences. Psychol Addict Behav. 2003; 17:328–31. [PubMed: 14640829]

30. Gough HG. Theory, development, and interpretation of the CPI Socialization scale. Psychol Rep. 1994; 75:651–700. [PubMed: 7809335]

31. Gough, HG.; Bradley, P. California Personality Inventory Manual. 3rd. Palo Alto, CA: Consulting Psychologists Press; 1996.

32. Cooney NL, Kadden RM, Litt MD. A comparison of methods for assessing sociopathy in male and female alcoholics. J Stud Alcohol. 1990; 51:42–8. [PubMed: 2299848]

33. Alterman AI, Rutherford MJ, Cacciola JS, McKay JR, Boardman CR. Prediction of seven month methadone maintenance treatment response by four measures of antisociality. Drug Alcohol Depend. 1998; 49:217–23. [PubMed: 9571386]

34. Kadden RM, Cooney NL, Getter H, Litt MD. Matching alcoholics to coping skills or interactional therapies: postreatment results. J Consult Clin Psychol. 1989; 57:698–704. [PubMed: 2557364]

35. Hattie J. Methodology review: assessing unidimensionality of tests and items. Appl Psych Meas. 1985; 9:139–64.

36. Hambleton RK, Rovinelli RJ. Assessing the dimensionality of a set of test items. Appl Psych Meas. 1986; 10:287–302.

37. McDonald RP, Ahlawat KS. Difficulty factors in binary data. Br J Math Stat Psychol. 1974; 27:82–99.

38. Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.

39. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In: Lord, FM.; Novick, MR., editors. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968. p. 395-479.

40. McDonald RP. A basis for multidimensional item response theory. Appl Psych Meas. 2000; 24:99–114.

41. McLeod, LD.; Swygert, KA.; Thissen, D. Test Scoring. Thissen, D.; Wainer, H., editors. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 189-216.

42. Baker, FB.; Kim, SH. Revised and Expanded. Second. New York, NY: Marcel Dekker; 1994. Item Response Theory: Parameter Estimation Techniques.

43. DeMars, CE. "Guessing" Parameter Estimates for Multidimensional IRT Models. Paper presented at the annual meeting the American Educational Research Association; Montreal. 2005.

44. Finch H, Habing B. Comparison of NOHARM and DETECT in item cluster recovery: counting dimensions and allocating items. J Educ Meas. 2005; 42:149–69.

45. Stout W, Habing B, Douglas J, Kim HR, Roussos L, Zhang J. Conditional covariance-based nonparametric multidimensionality assessment. Appl Psychol Meas. 1996; 20:331–54.

46. Mokken, RJ. A theory and procedure of scale analysis. The Hague: De Gruyter Mouton; 1971.

47. Zhang J, Stout W. Conditional covariance structure of generalized compensatory multidimensional items. Psychometrika. 1999; 64:129–52.

48. Sijtsma, K.; Molenaar, IW. Introduction to Nonparametric Item Response Theory. Thousand Oaks, CA: SAGE Publications; 2002.

49. Nunnally, JC. Psychometric Theory. 2nd. New York, NY: McGraw-Hill; 1978.

50. Thorndike, RL. Applied psychometrics. Boston: Houghton Mifflin; 1982.

51. Muthen, LK.; Muthen, BO. Mplus User's Guide. Los Angeles, CA: Muthen & Muthen; 1998.

52. Yu, CY. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes [PhD dissertation]. Los Angeles: University of California; 2002.

53. DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: considerations for the applied researcher. Pract Assess Res Eval. 2009; 14:1–11.

54. Dickenson, TS. Comparison of various ability estimates to the composite ability best measured by the total test score [PhD dissertation]. Columbia: University of South Carolina; 2005.

55. Cutrona, CE.; Russell, D. The provisions of social relationships and adaptation to stress. In: Jones, WH.; Perlman, D., editors. Advances in Personal Relationships. Greenwich, CT: JAI Press; 1987. p. 37-67.

56. Rush, BR.; Castel, S.; Brands, B.; Toneatto, T. Validation and comparison of screening tools for mental disorders in substance abusers. College on Problems of Drug Dependence; San Juan, Puerto Rico: 2008.

57. Kessler FHP, Cacciola JS, Faller ML, Formigoni MS, Cruz S, Brasiliano PF. Adaptação transcultural multicêntrica da sexta versão da Escala de Gravidade de Dependência (ASI6) para o Brasil. Rev Psiquiatr Rio Gd Sul. 2007; 29:335–6.

58. Díaz-Mesa EM, García-Portilla P, Sáiz P, Bobes MT, Casares MJ, Fonseca E, et al. Rendimiento psicométrico de la 6ª versión del Addiction Severity Index en español (ASI-6). Psicothema. 2010; 22:513–19. [PubMed: 20667284]

Table 1

Sample Characteristics

Variable	Psychometric Sample (N=607) M ± SD / % (n)	Validity Sample (subset, n=252) M ± SD / % (n)
Gender: Male	70.7 (429)	82.1 (207)
Race: African American	59.5 (361)	56.7 (143)
Caucasian	36.7 (223)	40.5 (102)
Other	3.8 (23)	2.8 (7)
Ethnicity: Hispanic	6.3 (38)	6.3 (16)
Age (years)	40.12 ± 10.72	41.04 ± 11.24
Education (years)	11.25 ± 2.04	11.22 ± 2.13
Worked for pay - past 30 days	22.6 (137)	23.0 (58)
Currently married or living as married	10.2 (62)	10.7 (27)
Living with children past 30 days	17.1 (104)	16.7 (42)
# prior substance abuse treatments	5.10 ± 5.36	5.75 ± 5.56
Years regular alcohol use	10.89 ± 11.32	10.81 ± 11.66
Days alcohol use - past 30 days	5.93 ± 8.70	4.00 ± 7.37
Years regular drug use	15.37 ± 9.90	15.43 ± 10.57
Days drug use - past 30 days	10.45 ± 10.63	9.03 ± 10.48
Primary Substance of Abuse: Alcohol	18.0 (109)	17.9 (45)
Cocaine	46.0 (279)	32.9 (83)
Heroin	23.7 (144)	36.1 (91)
Legally mandated to substance abuse treatment	19.4 (118)	26.6 (67)
History of arrest	88.1 (535)	89.7 (226)
Inpatient psychiatric treatment - lifetime	34.4 (209)	38.9 (98)
Days psychological problems - past 30 days	13.84 ± 12.16	13.10 ± 11.87
# medical hospitalizations - lifetime	3.91 ± 9.14	3.17 ± 6.21
Days physical problems - past 30 days	12.01 ± 12.17	12.37 ± 12.39

Table 2

ASI-6 Recent Status Scales, Items, and Psychometric Characteristics (n=607)

Scales	Items
Medical	
	Days in medical hospital - past 30 days
	Rating of physical health - past 30 days
	Days physical or medical symptoms or problems - past 30 days
	Days unable to carry out activities due to medical symptoms or problems - past 30 days
	Rating of severity of physical pain or discomfort - past 30 days
	Rating of worry/concern about physical health or medical problems - past 30 days
	Rating of current importance of treatment for physical health or medical problems
H = 0.56; minH1 = 0.36; alpha = 0.82	
Employment	
	Currently working
	Days worked for pay - past 30 days
	Money earned - past 30 days
	Days work related problems - past 30 days
H = 0.87; minH1 = 0.70; alpha = 0.89	
Alcohol	
	Days alcohol use - past 30 days
	Days ago, last drink
	Days of heavy drinking - past 30 days
	Money spent on alcohol - past 30 days
	Alcohol withdrawal - past 30 days
	Trouble controlling alcohol use - past 30 days
	Bothered by craving or urges to drink -past 30 days
	Various other problems due to drinking - past 30 days
	Days alcohol problems - past 30 days
	Rating of troubled/bothered by alcohol problems - past 30 days
	Rating of current importance of treatment for alcohol use
H = 0.71; minH1 = 0.60; alpha = 0.93	
Drug	
	Summed number of days of use for nine drug categories - past 30 days
	Number of different drug categories abused - past 30 days
	Days used drugs or abused prescribed medication - past 30 days
	Days ago, last used drugs of abused prescribed medications
	Money spent on drugs - past 30 days
	Drug withdrawal - past 30 days
	Trouble controlling drug use - past 30 days
	Bothered by craving or urges to use drugs - past 30 days
	Various problems because of drug use - past 30 days
	Days drug problems - past 30 days

Scales	Items
	Rating of troubled/bothered by drug problems - past 30 days
	Rating of current importance of treatment for drug use
H = 0.73; minH1 = 0.63; alpha = 0.94	
Legal	
	Illegal income - past 30 days
	Days sold or manufactured drugs - the past 30
	Days robbed, assaulted or threatened anyone - past 30 days
	Days illegal activity for profit or against property - past 30 days
	Days other illegal activities - past 30 days
	Days any illegal activity - past 30 days
H = 0.54; minH1 = 0.28; alpha = 0.78	
Family/Social Problems	
	Trouble getting along with partner, relatives or friends - past 30 days
	Had arguments with partner, relatives or friends - past 30 days
	Had interactions with partner, relatives or friends result in physical violence - past 30 days
	Rating of troubled/bothered by problems with adult relationships - past 30 days
	Rating of current importance of treatment for problems with adult relationships
H = 0.49; minH1 = 0.46; alpha = 0.71	
Family/Social Support	
	Had contact with partner, relatives or friends - past 30 days
	Spent time with partner, relatives or friends - past 30 days
	Talked about feelings or problems with partner, relatives or friends - past 30 days
	Can count on partner, relatives or friends - past 30 days
H = 0.54; minH1 = 0.40; alpha = 0.78	
Child Problems	
	Number of children with a serious problem - past 30 days
	Rating of how necessary current additional services are to treat children's' problems
	Rating of trouble getting along with children - past 30 days?
	Rating of current importance of counseling to help get along with children?
	Rating of current need for childcare
H = 0.51; minH1 = 0.35; alpha = 0.73	
Psychiatric	
	Days in psychiatric hospital - past 30
	Rating of troubled/bothered by symptoms related to trauma - past 30 days
	Rating of need for treatment related to trauma - past 30 days
	Trouble sleeping - past 30 days
	Depressed - past 30 days
	Anxious - past 30 days
	Hallucinations - past 30 days
	Trouble thinking, concentrating, remembering - past 30 days
	Difficulty controlling temper or violent urges - past 30 days
	Physically violent towards someone - past 30 days

Scales	Items
	Serious thoughts of suicide - past 30 days
	Attempted suicide - past 30 days
	Recency of serious psychological problems
	Days experienced psychological problems - past 30 days
	Days unable to carry out normal activities because of psychological problems - past 30 days
	Rating of troubled/bothered by psychological problems - past 30 days
	Rating of current importance of treatment psychological problems

H = 0.49; minH1 = 0.34; alpha = 0.87

Table 3

Comparison of the *H*, *minHᵢ*, and *α* values[*] for the entire sample versus the lowest values found for any of the four demographic subgroups

Scale	H	minHᵢ	α
Legal[a]	0.54 vs. 0.51	0.28 vs 0.23	0.78 vs. 0.75
Family/Social Problems	0.49 vs. 0.41	0.46 vs. 0.33	0.71 vs. 0.68
Child Problems	0.51 vs. 0.45	0.35 vs. 0.34	0.73 vs. 0.71
Psychiatric[b]	0.48 vs. 0.45	0.34 vs. 0.26	0.87 vs. 0.86

[*]
H = overall scalability coefficient; $minH_i$ = minimum pairwise scalability coefficient; and $α$ = Cronbach's alpha coefficient

[a]
two of the four subgroups had a negative $minH_{ij}$ (minimum item-pair scalability coefficient) value. The entire sample had a $minH_{ij}$ of 0.06.

[b]
Three of the four subgroups had a negative $minH_{ij}$ value, as did the scale for the entire sample

Table 4

Intercorrelations between the ASI-6 Recent Status Scales (n= 607)

Scale	M	E	A	D	L	FP	FS	CP	P
Medical (M)	1.00	.13	.19	.21	.02	.22	.01	.12	.44
Employment (E)		1.00	-.06	-.08	-.01	-.06	-.08	-.05	.06
Alcohol (A)			1.00	.03	.07	.22	-.02	.02	.30
Drug (D)				1.00	.40	.32	.10	.13	.33
Legal (L)					1.00	.24	.03	.07	.16
Family/Social Problems (FP)						1.00	.12	.22	.42
Family/Social Support (FS)							1.00	.19	-.08
Children Problems (CP)								1.00	.11
Psychiatric (P)									1.00

Table 5

Correlations between ASI-6 Recent Status Scales and External Validity Measures (n=252)

Scales	External Validity Measures*								
	SF-12 Physical	SAS-SR Work[b]	SIP	SIP-D	CPI-So	Prior arrests[a]	SAS-SR Social	SCL-10R	SF-12 Mental
Medical	**-.64**	.11	.15	.19	-.23	.00	.18	.44	-.34
Employment	-.13	**.76**	-.06	-.03	-.14	.14	.02	.00	.02
Alcohol	-.07	-.04	**.68**	.01	-.11	-.17	.21	.29	-.22
Drug	-.13	-.14	-.14	**.61**	-.17	.02	.18	.30	-.27
Legal	.08	-.06	-.04	.28	**-.14**	**.15**	.09	.09	-.09
Family/Social Problems	-.16	-.13	.17	.25	-.21	-.12	**.40**	.36	-.36
Family/Social Support	-.06	-.09	-.15	.00	.11	-.03	**-.34**	-.12	.01
Child Problems	.04	-.04	-.13	.12	-.07	-.12	**-.04**	.15	-.16
Psychiatric	-.24	.00	.26	.37	-.34	-.14	.44	**.68**	**-.61**

*
Higher scores on the SAS-SR, SIP, SIP-D, Prior arrests, and SCL-10R indicate more problems.

Higher scores on the SF-12 and CPI-So indicate fewer problems.

[a]
number of arrests in prior two years

[b]
employed vs. not employed

Bold indicates correlations between ASI-6 RSSs and their corresponding validity measure(s)