Addiction severity assessment tool: Development of a self-report measure for clients in substance abuse treatment

Stephen F. Butler a,∗, Simon H. Budman a, Michael D. McGee b, Michael Sean Davis a, Rebecca Cornelli a, Leslie C. Morey c

a Inflexxion Inc., 320 Needham St., Suite 100, Newton, MA 02464-1594, USA
b St. Joseph Hospital, Nashua NH 03060, USA
c Texas A&M University, College Station, TX 77843, USA

Received 11 May 2004; received in revised form 27 April 2005; accepted 6 May 2005

Abstract

This article describes the development and reliability and validity testing of the Addiction Severity Assessment Tool (ASA T), a brief, 27-item multidimensional self-report measure of problem severity in daily functioning, relational functioning, dysphoric states, dependence severity, recovery skill/self-efficacy, and existential factors for adult substance abuse clients. Items generated for an Alpha version were conceptually and empirically evaluated. A Beta version underwent further empirical evaluation and item selection. Cross validation of the final version examined internal consistency, test-retest reliability, factor structure, and convergent/discriminant and known groups validity. Sensitivity to change was evaluated in a 3-month outcome study. Clients were recruited from inpatient, outpatient and residential substance abuse treatment centers, and a sample of 238 nonpatients were also recruited from community groups. The Beta version was tested with 201 clients, and cross validation involved 242 clients. Well-known standardized, self-report and interview-based comparison measures were used to test convergent/discriminant validity of the ASA T. Reliability coefficients for the six ASA T domains were acceptable. Reasonable convergent/discriminant and known groups, construct validity were demonstrated, along with sensitivity to change of the domain scale scores. The ASA T appears to comprise a useful new tool for assessing clinical outcomes of adult clients in substance abuse treatment.

© 2005 Elsevier Ireland Ltd. All rights reserved.

Keywords: Drug and alcohol assessment; Self-report; Problem severity; Brief assessment; Addiction

1. Introduction

The ASA T is a self-report measure of problem severity in adult substance abuse clients that is easy to administer and score, multidimensional, and psychometrically sound. This manuscript describes the creation, testing, and cross-validation of the measure. The ASA T contains 27 items and is rated on a four-point scale from “not at all true” to “very true,” addressing the following subscales: daily functioning, relational functioning, dysphoric states, dependence severity, recovery skill/self-efficacy, and existential factors. Increasing scores denote increasing severity.

Development of the ASA T was stimulated by the need in clinical substance abuse settings for a brief, easy to use, validated, self-report, and multidimensional assessment that provides information relevant for treatment planning and outcome measurement (Sederer et al., 1996). Given the realities of most clinical settings, outcome studies typically require a single self-report assessment tool. In addition to being reliable and valid, such a tool must be easily and quickly administered and scored and must tap outcome in the multidimensional areas of symptoms and functioning relevant to substance abuse populations. Comparable measures for mental health and medical populations include the BASIS-32 (Eisen et al., 1986; Eisen, 2000) and the Short-Form-36 (SF-36; Ware et al., 1993). These short, multidimensional
measures are widely used. However, no such measure is available that is specifically developed for clients in substance abuse treatment.

Self-report tools for adult populations do exist in the substance abuse field. While a comprehensive review is beyond the scope of this article, there are some well-known assessments. None of these measures provide an overall view of adult substance abuse clients’ problem severity while also being brief and easy to administer. For instance, the Drug Abuse Screening Test (DAST; Skinner, 1982) is primarily symptom-focused and does not reflect the multidimensional nature of addiction problems. The Recovery Attitude and Treatment Evaluation-Questionnaire I (RAA TE-QI; Smith et al., 1995) is multidimensional, self-report, and was developed for addicted populations. It has some published psychometrics, but it is not brief (94 items), and there are no published data regarding the RAA TE-QI’s sensitivity to change. This is probably because the focus of the measure is on the client’s readiness for treatment; however, this makes the scale less useful for outcome studies. The Substance Abuse Subtle Screening Inventory (SASSI; Miller, 1985) is another multidimensional self-report screening measure with some validity and a small amount of reliability data published in its manual. Its focus is to “identify individuals who have a high probability of having a substance use disorder” (SASSI Institute, 2003), which limits its utility for assessing adults in treatment for substance abuse. The Drug Use Screening Inventory (DUSI) is multidimensional and has been shown to be sensitive to change. It is long for a self-report measure (159 items). Although there is an adult version, the reliability and validity data that are published (Tarter and Hegedus, 1991; Tarter and Kirisci, 2001) or available at their website (Gordian Group, 2003) are for adolescent populations only.

Probably the best known and most well-researched multidimensional assessment designed specifically for adult substance abuse populations is a structured clinical interview called the Addiction Severity Index (McLellan et al., 1985). This interview is particularly useful for gathering detailed information consistent with a psychosocial report. It also includes summary composite scores calculated for each of seven problem areas (medical, employment, alcohol, drug, legal, family, and social, and psychiatric). These composite scores are recommended for research analyses (McLellan et al., 1985). A major drawback of the ASI has been its lengthy interview format and the requirement of specialized training for interviewers. Many of these concerns have been addressed with a newly developed, reliable and valid multimedia computer-administered version (ASI-MV; Butler et al., 2001). However, despite the solutions offered by the ASI-MV, the interview still takes about 45 min to complete and may not be suitable for many purposes.

While the assessments reviewed above are well-known, literature reviews and reviews of compilations of substance abuse measures (e.g., Allen and Columbus, 1995) failed to yield any brief, multidimensional self-report assessment of addiction-related problems. We, therefore, set out to develop a practical and psychometrically sound assessment that would yield a score for problem areas similar to the BASIS-32 and the SF-36 but relevant to adult substance abuse clients in treatment. The principles guiding scale development were that the ASAT should: be self-report (6th grade reading level or less), provide an assessment of the severity of addiction-related problems, be brief enough to be used in typical clinical settings, be easy to administer and score, assess multidimensional domains of symptoms and functioning affected by substance abuse, be psychometrically sound, and be sensitive to change so that it is capable of detecting the effects of treatment over time. These characteristics compare well to those described by Carroll and Rounsaville (2002) as essential qualities of assessments in the field of substance abuse treatment.

2. Methods

Our scale development procedure was based on classical methods for establishing content validity of a scale (e.g., Anastasi and Urbina, 1997). Development of the ASAT occurred in three stages to ensure that the resulting measure would be reliable and demonstrate discriminant validity and sensitivity to change. In the first stage, content valid domains and items were generated based on literature reviews and consulting with experts. This resulted in an initial “Alpha version” of the ASAT. The Alpha ASAT was tested on addicted clients and a community sample for item evaluation and reduction. This resulted in an 83 item “Beta version” of the ASAT. The Beta version was tested on addiction clients in inpatient, outpatient, and residential settings and subjected to further, extensive item evaluation and item reduction. The resulting final version of the ASAT was then cross validated in the third stage. This article presents a summary of the first two stages and presents in detail the methods and results of the cross validation. Result summaries of the first two stages can be found online (Appendix A).

2.1. Content validity: domain and item development

The primary focus of the ASAT was determined at the outset to follow the recommendations of prominent addiction assessment theorists (e.g., McLellan and Durell, 1996; Tonigan, 1995) that assessment of substance abuse treatment outcome should be multidimensional. These authors and others (e.g., Allen and Columbus, 1995) have promoted the idea that outcome assessments in addictions should tap, in addition to substance use, life-functioning areas affected by, and possibly influencing one’s substance use. The first step in developing the ASAT, therefore, was to determine what dimensions of functioning it should assess. A literature review suggested that a measure like the ASAT should evaluate well-being, psychological status, functional ability (Sedere et al., 1996), improvements in health, personal, and social functioning (McLellan and Durell, 1996; McLellan
et al., 1985, 1992), and readiness to change (Prochaska et al., 1992; Miller and Rollnick, 1991). Reviews of other substance abuse measures such as the DUSI, RAATE-QI, and the ASI supported this general list. Thus, six initial domains of functioning and health status were identified: daily living skills, relation to self and others, psychiatric impairment, problem acceptance and change readiness, dependence severity, and recovery stresses.

The next step in delineating and defining the domains of interest was to consult outside experts in the field of substance abuse. An important goal was to include nationally prominent addictions researchers, many of whom are also practicing clinicians. We also sought the input of well-known clinicians whose primary focus is direct service, rather than research. A list of 23 individuals was created that balanced researchers of national standing with front-line clinicians based on our knowledge of the literature and of the clinical community. These 23 individuals were approached to be interviewed, and 19 agreed. Of the 19, nine were researchers, or researchers and clinicians, and 10 held primarily clinical (and often administrative) positions. Eight held M.D. degrees, eight were Ph.D.s, and three had masters degrees. The experts’ task was to: (1) consider the domains and definitions we had constructed; (2) decide if these are the most important dimensions for inclusion in an ideal outcome measure; (3) decide if there are other dimensions that should be included; and (4) determine whether the domains are distinct. Feedback from phone interviews with these experts resulted in a revised list of eight domains: daily functioning, relational functioning, dysphoric states, problem acceptance/change readiness, dependence severity, recovery skills/self-efficacy, and existential factors. The latter domain, existential factors, reflected the input of three experts (all of whom were direct service providers) that the ASAT should acknowledge the importance of spiritual concerns in the role of recovery without necessarily endorsing a 12-step philosophy. Once the initial domains were identified, definitions of the domains were created and reviewed by the experts. The research team integrated these suggestions into final definitions for the domains.

Using the domains and definitions constructed as a result of the expert survey, the research team created a pool of 742 self-report items corresponding to the identified domains. Then the authors rated each item on two criteria: the domain the item corresponded to, and the quality of the item on a 5 point scale (1 being the highest quality). Items that had less than a 60% agreement as to which domain they pertained to, or a mean quality rating greater than 3 were deleted. This reduced the item pool to 290 items. Next, the items underwent further reduction by seven independent judges who were substance abuse experts not affiliated with this project. These judges repeated the classification and quality rating procedure described above. Only items that achieved a 70% agreement in domain classification, and a mean quality score below 3 were retained. A Flesch-Kuncaid reading level was obtained for each item. Any item that exceeded a 6th grade reading level was eliminated. These procedures resulted in 133 items which comprised the initial “Alpha version” of the ASAT. All items were in the form of a simple, declarative statement about the person (e.g., “I have too much free time.”), which are rated on a four-point scale from 1 = “Not at All True” to 4 = “Very True.”

2.2. Empirical evaluation of the Alpha version

Next, the Alpha ASAT was administered to participants. These data were used to further reduce the number of assessment items using the following criteria: (1) item distributions should be reasonably variable; (2) items should discriminate non-patient, former patient (aftercare), and current patient populations; (3) items should be consistent, but not redundant, with other items from the same domain to maximize the efficiency of the scale in assessing the domain under consideration; (4) items should not be overly sensitive to social desirability; (5) items should be domain specific; and (6) items should have good test-retest reliability.

Two-hundred-forty-three participants were recruited (86 inpatients, 54 aftercare patients—of which 39 were followed for 1 week to examine test-retest, and 103 non-patients from the community). Patient participants meeting inclusion/exclusion criteria were identified by clinic staff and approached. Of these, approximately 86% agreed to participate. Participants were administered the Alpha ASAT and a short form of the Marlowe-Crowne Social Desirability Scale (Reynolds, 1982) which assesses a person’s proclivity to answer in a socially desirable way.

All items had acceptable variability, encompassing the entire response range from 1–4. Items generally were able to discriminate between patient and community populations using a one way ANOVA with a post-hoc Tukey’s t-test. No items were eliminated based on acceptable variability or ability to discriminate the participant populations. Item consistency was examined by calculating a corrected item total correlation (sometimes called the part-whole correlation) for each item. This procedure involves correlating each item with the sum of the other items in a particular domain. This is a test for redundancy, so the correlation should be high, suggesting that it measures the same construct, but not too high, which would suggest that the item is redundant. As an adjunct to the corrected item total correlation, the squared multiple correlation between each item and all other items from the same scale was calculated. These values were useful in identifying highly redundant items, since the response to a redundant item should be easily predicted from the other items in the scale. Item social desirability was assessed by checking to make sure that each item correlated higher with its parent domain than it did with the Marlowe-Crowne social desirability scale. Item domain specificity was examined; items that demonstrated greater correlations with domains other than their parent domain were deleted. Test-retest reliability was also examined for each item. Generally items were deleted if their test-retest correlation was below 0.50, although in several cases exceptions were made to retain a large enough
number of items in each domain. As a result of this item analyses, 50 items were eliminated resulting in an 83-item “Beta version” of the ASA T.

2.3. Empirical evaluation of the Beta version

In order to further reduce the number of items, the Beta version was tested in a substance abuse population. Items on the Beta version were examined for their contribution to the internal consistency of subscales, test-retest reliability, criterion validity, and sensitivity to change.

2.3.1. Participants

Another 201 clients in treatment for substance abuse disorders were recruited from inpatient, outpatient and residential, public and private treatment centers in New England. Patient participants meeting inclusion/exclusion criteria were identified by clinic staff and approached to participate. About 80% agreed to sign consent. The average age of the clients was 36.5 years (S.D. = 8.3), 41% were women, and 37% were minority (African American, Hispanic, and Native American). Approximately half \( N = 101 \) were clients newly admitted to treatment (within 2 weeks) in order to permit testing of item sensitivity to change, while the other half \( N = 100 \) were “stabilized” clients who had been in treatment at least 2 weeks. The stabilized clients allowed us to assume a certain degree of stability in order to evaluate test-retest reliability of items. Fifty of the stabilized clients were randomly selected to participate in the evaluation of test-retest reliability, and 67 of the newly admitted clients were selected to be followed at 3 months to examine sensitivity to change.

2.3.2. Measures

All participants were administered the Beta ASAT along with a battery of self-report comparison measures to assess item construct validity. We administered the addiction severity index (ASI; McLellan et al., 1985), which was described previously. A combination of the Alcohol and Drug domain composite scores of the ASI was used as the convergent marker for dependence severity items. The Social Adjustment Scale-Self Report (SAS-SR; Weissman and Bothwell, 1976) is a 54-item self-report scale for measuring instrumental or expressive role performance in areas of functioning. The social and leisure subscale was the convergent marker for the Daily Functioning domain and the Total SAS score was examined to determine which were most sensitive to change. The DASES was used to evaluate recovery skill/self-efficacy items.

The DASES was used to evaluate recovery skill/self-efficacy items. The Purpose In Life (PIL; Crumbaugh and Maholick, 1964) is a 20-item self-report attitude scale designed to measure the failure to find meaning and purpose in life (Frankl, 1972). The PIL total score was the convergent marker for existential factors items. Finally, the Stages of Change-Readiness and Treatment Eagerness Scale (SOCRATES; Miller and Tonigan, 1996) was the convergent marker for the problem acceptance/change readiness items.

2.3.3. Item evaluation

Item-level analyses were again conducted with respect to item total correlation, squared multiple correlation, and 1-week item-level test-retest reliability. In addition, criterion validity (correlation with its comparison measure) was examined, along with each item’s correlation with other items in its domain, as well as the highest other domain correlation. Correlations with the Marlows-Crown Social Desirability scale were not examined at this stage. Criterion validity was evaluated by examining correlations with an item’s comparison scale. In addition, each item’s correlation with the other items in its domain was examined, along with the highest correlation with other domain. Finally, 67 participants in active treatment were randomly selected from the larger pool of clients in treatment to be followed for 3 months. Of these, 47 (70%) were successfully followed 3 months later. Items were examined to determine which were most sensitive to change. Thus, the ideal item was: (1) non-redundant; (2) consistent with the other items in its domain; (3) correlated highly with its comparison measure; (4) correlated higher with its own domain than any other domain; (5) test-retest reliability greater than 0.50; and (6) demonstrated significant change at follow-up.

At this stage, it became evident that the problem acceptance/change readiness domain was exceedingly context-dependent. While an individual’s answer to these items makes sense on intake into a treatment program, at follow-up, interpreting a client’s answer to a question like “I don’t really want to make any changes in my drinking or drug use right now,” depends on the context of that person’s situation. Because of this and consistently poorer performance of this domain on the item parameters \( \alpha = 0.50 \), low values for corrected item total correlation, squared multiple correlation, and mean change), the problem acceptance/change readiness domain was deleted and is not considered further.

Items with the best balance of item parameters were retained for the remaining six domains of a 27-item version of the ASAT. Table 1 presents the ASAT domains, the number of items comprising each domain, a brief definition of the domain, and sample items. The Drug Avoidance Self-Efficacy Scale (DASES; Martin et al., 1995) is a 16-item self-report questionnaire assessing clients’ self-efficacy to resist drug or alcohol use in different situations in which they might be tempted to use.

---

2 Summaries of the item analyses are presented in the online supplement to this paper at http://dx.doi.org.
Table 1  
ASA T subscale or domain names, the number of items comprising each domain, a brief definition of high scores of each the domain, and sample items.

<table>
<thead>
<tr>
<th>ASA T domain</th>
<th>No. of items</th>
<th>Meaning of high scores</th>
<th>Sample item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily functioning</td>
<td>5</td>
<td>Difficulty achieving a “balance” between work and leisure. Leisure may be limited to substance-related activities.</td>
<td>I am able to do the things I need to do each day (R).</td>
</tr>
<tr>
<td>Relational functioning</td>
<td>3</td>
<td>The quality of relationships is characterized by interpersonal distance. Important relationships may be unsupportive of sobriety.</td>
<td>My close friends are there for me when I need them (R).</td>
</tr>
<tr>
<td>Dysphoric states</td>
<td>4</td>
<td>Presence of difficult-to-manage, negative, dysphoric emotions, including anxiety, depression, or irritability.</td>
<td>I feel angry a lot.</td>
</tr>
<tr>
<td>Dependence severity</td>
<td>5</td>
<td>Presence of psychological and social aspects of dependence, including compulsive substance use, loss of control, use of substances despite adverse consequences, and use of substances to cope.</td>
<td>I can’t ignore my cravings for alcohol or drugs.</td>
</tr>
<tr>
<td>Recovery skills/efficacy</td>
<td>5</td>
<td>Lacking attitudes, beliefs and skills that could facilitate staying clean in the presence of stressful events.</td>
<td>I don’t need alcohol or drugs to have fun (R).</td>
</tr>
<tr>
<td>Existential factors</td>
<td>5</td>
<td>Failure to endorse a subjective experience of purpose, meaning, or spiritual center in one’s life.</td>
<td>I have goals in my life (R).</td>
</tr>
<tr>
<td>Total score</td>
<td>27</td>
<td>Overall problem severity.</td>
<td></td>
</tr>
</tbody>
</table>

*R: reverse scored item.

3. Cross validation of the ASA T

3.1. Participants

Two-hundred-forty-two (242) individuals in substance abuse treatment were recruited to participate in the cross validation of the ASA T from inpatient, outpatient and residential, public and private programs in New York, Massachusetts, Rhode Island, and Florida. As in the other stages, patient participants meeting inclusion/exclusion criteria were identified by clinic staff and approached. Of these, approximately 91% agreed to participate. The average age of clients was 38.4 years (S.D. = 7.9), 38% were women, and 25% were married, 34% never married, and the remaining 41% were divorced, separated or widowed. The racial breakdown of the sample was as follows: 57% white, 25% African American, 16% Hispanic, and 2% Asian and Native American. These sample characteristics compare well with national samples of clients in substance abuse treatment in the US (women 33%, African American 17%, Hispanic 11%, and Asian/Native American 2%; SAMHSA, 2002). Thirty percent of the sample had less than a high school education and 11% reported completing middle school or less education. Finally, the substances of abuse reported by subjects included alcohol (56%), heroin (35%), and amphetamines (12%). Seventy-four percent (74%) reporting using more than one substance as their primary problem.

An additional sample of 238 non-patient, community participants was recruited. Social organizations such as churches,Parent Teacher Associations, Lions Club, etc. were contacted to administer forms to their members in exchange for a contribution to their organization. Seven organizations were so identified. The average age of this sample was 41.8 years (S.D. = 16.5). 43.5% were women, 54.9% were married, 30.5% single, and 14.6% were divorced, separated or widowed. Most (64.1%) were white, 15.6% were African American, 16.0% were Hispanic, and 4.2% were Asian or other.

3.2. Measures

Client participants were administered the 27 item ASA T along with the battery of comparison measures described above for testing the Beta version of the scale. Community participants were also administered the 27 item ASA T. Since the community sample contained individuals who were not identified as having substance abuse problems, it was important to determine which of these individuals might have problems with alcohol or drugs. Thus, these participants were administered the alcohol use disorders identification test (AUDIT; Saunders et al., 1993) and the DAST (Skinner, 1982). These well-known screening devices permitted us to estimate the level of substance abuse problems within the community sample.

3.3. Procedures

The procedures for cross validation were similar to those described above for the Beta version. All participants were administered the ASA T and the comparison measures. The sample was structured so that 142 were clients newly admitted to treatment (within 2 weeks) and 100 were “stabilized” clients who had been in treatment at least 2 weeks. From those newly admitted to treatment, 111 were randomly chosen to test sensitivity of the ASA T to change in a 3-month follow-up study. At the follow-up assessment, these participants were administered the ASA T and the follow-up version of the ASI interview. Of the stabilized participants, 50 were...
randomly selected to evaluate test-retest reliability and were re-administered the ASAT 1-week following the baseline administration.

For the community sample, participants completed their assessment battery in small groups (approximately 10 at a time). A contact person at the participating organization described the purpose of the study from a provided script. The assessment battery was then passed out, completed and returned to the research team.

4. Results

Below we present findings from the cross validation study of the 27 items ASAT.

4.1. Reliability

4.1.1. Internal consistency

Coefficient alpha was calculated for each of the six ASAT domains based on data from the 242 client participants. Although adequate coefficient alpha is more difficult to obtain from brief scales than it is for long scales, all ASAT domains achieved coefficient alphas greater than 0.70. Specifically, the coefficient alphas were as follows: daily functioning (five items) $\alpha = 0.79$, relational functioning (three items) $\alpha = 0.78$, dysphoric states (four items) $\alpha = 0.79$, dependence severity (five items) $\alpha = 0.80$, recovery skill/self-efficacy (five items) $\alpha = 0.81$, and existential factors (five items) $\alpha = 0.81$. Coefficient alpha was calculated for a total score (average of all items) and achieved an acceptable level $\alpha = 0.93$.

Some authors (e.g., Nunnally and Bernstein, 1994) suggest that clinical measures should have alphas in excess of 0.90. Others (e.g., Steiner, 2003a,b) persuasively argue that internal consistencies this high most likely represent unnecessary redundancy. In some cases (e.g., Loevinger, 1954), increasing coefficient alphas can actually lower validity. To obtain a very high alpha value, the ASAT would need more items, which would defeat our goal of creating an established outcome assessment, the ASI. Only three of the ASAT domains had alphas greater than 0.80 or higher. Three domains achieved large effects: dependence severity, recovery skill/self-efficacy, and daily functioning (1.17, 0.83, and 0.83, respectively), which may be the areas most likely targeted by substance abuse treatment. Effect sizes between small (0.20) and medium (0.50; Cohen, 1988) were observed for the relational functioning (0.34), and existential factors (0.28) subscales. Larger changes in relational functioning may require more intensive intervention (e.g., psychotherapy), although the effect size for dysphoric states subscale was large (0.61). As expected, the effect size for the total score was large at 0.88.

4.1.3. Sensitivity to change

Sensitivity to change was evaluated in two ways. First, to examine the extent to which clients show statistically significant change on each domain, paired t-tests compared Time 1 with Time 2 follow-up domain scores. Second, in an assessment of whether another, established measure would show change, we examined change as measured by ASI composite scores and related this change to that obtained on relevant domains of the ASAT.

At follow-up, we were able to contact 70% ($N = 78$) of the 111 randomly selected to be followed. Of those, one (0.9%) declined to participate, one was in prison and unable to participate, and one was dead. The remaining 75 individuals (66% of the sample) participated in the follow-up procedures. For outcome studies involving adult substance abusers, which often are difficult to follow due to unstable lifestyles of some individuals, McLellan and Durell (1996) recommend follow-up rates of 70% or better. Analyses were conducted to identify any systematic differences between those successfully followed (completers) and non-completers. The variables examined were age, gender, ethnicity (white versus minority), employment status (employed versus unemployed), severity of problems as measured by the ASI composite scores for medical, employment, alcohol, drug, legal, family and social, and psychiatric problems, education (less than high school versus high school or greater), and marital status. The two groups did not differ significantly on any of these variables. Thus, it is reasonable to assume that the follow-up sample is representative of the recruited sample.

Paired t-tests were run for all ASAT domains. The tests revealed significant improvement (i.e., lower group means) for the total score was large at 0.88. A second set of analyses examined correspondence of the change measured by the ASAT with that measured by an established outcome assessment, the ASI. Only three of the ASI domains were thought to correspond to the ASAT domains and the content focuses of even these domains are different. Nevertheless, for those domains that do have a corresponding ASI subscale, a reasonably high correspondence should be expected. Therefore, we compared change on the ASI alcohol and drug composite scores to change on the dependence severity and recovery skill/self-efficacy domains, change on the ASI Family composite with change...
The goodness of fit index (GFI; Hu and Bentler, 1995) was only values greater than 0.90 are considered acceptable. The comparative fit index (CFI; Bentler, 1990) was acceptable (CFI = 0.93; generally an examination of fit indices is in order. The fit indices revealed a mixed picture. The model yielded a chi-square of 501.1, d.f. = 302, which was not significant at the Bonferroni-corrected alpha level. Although non-significant chi-squares are desirable, they are not expected to be related than they do with measures to which they are expected to be related. In this study, participants who took the ASAT also completed established comparison measures considered to correspond to the domain concepts assessed by the ASAT. It should be noted that it was assumed that there would be substantial intercorrelation among the ASAT domains and across the comparison measures. While it is important to distinguish the various ASAT domains, it is unlikely that “daily functioning,” for instance, will be statistically independent of “relational functioning” or “dysphoric states.” Clinically and phenomenologically, problems in such areas tend to co-occur.

Another issue that complicates examinations of convergent and discriminant validity is the degree of discriminate validity of the comparison measures. While we chose the best

### Underlying factor structure

To examine the underlying factor structure of the ASAT we used procedures for confirmatory factor analysis (CFA; Bentler and Stein, 1992). The ASAT’s factorial validity model (first-order CFA model; Byrne, 2001) was created based on data collected during testing of the Beta version of the ASAT. A confirmatory analysis was then conducted on this model with the cross-validation sample. The first-order CFA model treats each ASAT domain as a factor (latent variable) that is predicted by that domain’s items (observed variables). The model further stipulates that the ASAT domains (latent variables) are not independent (i.e., are likely to be correlated). The model yielded a chi-square of 501.1, d.f. = 302, p < 0.001. Although non-significant chi-squares are desirable, it is common for models to achieve significance, so generally an examination of fit indices is in order. The fit indices revealed a mixed picture. The comparative fit index (CFI; Bentler, 1990) was acceptable (CFI = 0.93; generally values greater than 0.90 are considered acceptable). The goodness of fit index (GFI; Hu and Bentler, 1995) was only 0.86, which was below the recommended 0.90. Examination of the root mean square error of approximation (RMSEA) statistic provides information about the covariance structure of the model and in this case was an acceptable 0.055 (values between 0.05 and 0.08 reflect a reasonably good fit, Byrne, 2001). Finally, examination of the factor loading estimates for the final model (domain relevant items loading on their domain) revealed that these were reasonable and statistically significant. The factor loading estimates for the ASAT domains are: daily functioning = 0.68, range 0.45–0.86; relational functioning = 0.74, range 0.61–0.82; dysphoric states = 0.71, range 0.60–0.80; dependence severity = 0.70, range 0.39–0.88; recovery skill/self-efficacy = 0.63, range 0.35–0.82; and existential factors = 0.67, range 0.53–0.83. Taken together, these results suggest that given the conceptual domain construction method used to create the ASAT, the measure confirms a reasonable factor structure.

### Convergent and discriminant validity

The ASAT was to be subjected to tests of convergent and discriminant validity (Campbell and Fiske, 1959). In general convergent validity focuses on the extent to which scales measuring the same construct correlate with each other. Conversely, discriminant validity demonstrates that the measure does not correlate with measures that are intended to be different. As Stewart et al. (1993, p. 315) note, “the appropriate test is whether measures correlate lower with measures to which they are not expected to be related than they do with measures to which they are expected to be related.” In this study, participants who took the ASAT also completed established comparison measures considered to correspond to the domain concepts assessed by the ASAT. It should be noted that it was assumed that there would be substantial intercorrelation among the ASAT domains and across the comparison measures. While it is important to distinguish the various ASAT domains, it is unlikely that “daily functioning,” for instance, will be statistically independent of “relational functioning” or “dysphoric states.” Clinically and phenomenologically, problems in such areas tend to co-occur.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Baseline mean (S.D.)</th>
<th>Follow up mean (S.D.)</th>
<th>d.f.</th>
<th>t</th>
<th>p</th>
<th>Effect size</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily functioning</td>
<td>2.5 (0.67)</td>
<td>1.9 (0.76)</td>
<td>74</td>
<td>6.9</td>
<td>&lt;0.001</td>
<td>0.83</td>
<td>0.45-0.78</td>
</tr>
<tr>
<td>Relational functioning</td>
<td>2.4 (0.85)</td>
<td>2.1 (0.91)</td>
<td>74</td>
<td>2.4</td>
<td>0.019</td>
<td>0.34</td>
<td>0.05-0.57</td>
</tr>
<tr>
<td>Dysphoric states</td>
<td>2.5 (0.73)</td>
<td>2.0 (0.90)</td>
<td>74</td>
<td>3.8</td>
<td>&lt;0.001</td>
<td>0.61</td>
<td>0.22-0.66</td>
</tr>
<tr>
<td>Dependence severity</td>
<td>2.8 (0.88)</td>
<td>1.8 (0.82)</td>
<td>74</td>
<td>8.6</td>
<td>&lt;0.001</td>
<td>1.17</td>
<td>0.77-2.10</td>
</tr>
<tr>
<td>Recovery skill/efficacy</td>
<td>3.4 (0.79)</td>
<td>2.4 (0.94)</td>
<td>74</td>
<td>5.9</td>
<td>&lt;0.001</td>
<td>0.83</td>
<td>0.36-0.75</td>
</tr>
<tr>
<td>Existential factors</td>
<td>2.0 (0.70)</td>
<td>1.8 (0.73)</td>
<td>74</td>
<td>3.3</td>
<td>0.002</td>
<td>0.28</td>
<td>0.09-0.38</td>
</tr>
<tr>
<td>Total average</td>
<td>2.4 (0.50)</td>
<td>1.9 (0.63)</td>
<td>74</td>
<td>7.2</td>
<td>&lt;0.001</td>
<td>0.88</td>
<td>0.39-0.69</td>
</tr>
</tbody>
</table>

N for all comparisons = 75.
Table 3 presents the matrix of correlations between the ASA T domain scores (initial administration for all participants) with comparison measures (convergent indicators).

Table 4 presents the results of univariate analysis of variance for each ASA T domain score as the dependent variable and participant group as the fixed factor at four levels (in treatment, aftercare, community with substance problems, and community with no substance problems). As expected, for most of the domains and the total score, the treatment group had greater scores than the aftercare group, which, in turn, had greater scores than the community sample with no known substance issues. For four of the six domains (daily functioning, relational functioning, dysphoric states, and dependence severity), the 30 individuals in the community sample had scores that were not different from those obtained by clients in aftercare. Although it is premature to draw definitive conclusions from this small sample of individuals in the community who have screened for substance problems, it is interesting to note that their scores on recovery skill/self-efficacy were similar to the scores obtained by clients in treatment and were not different from those obtained by clients in aftercare.

A final test of construct validity, referred to as known groups validity (Anastasi and Urbina, 1997; Stewart et al., 1993), was conducted. This type of validity test requires that the assessment differentiate groups known to be different in expected ways. Four groups were created: clients in treatment, clients in aftercare, community participants whose scores on the AUDIT and/or DAST suggested they may have substance abuse problems, and community participants with no known substance problems. In general, we expected those in treatment to have the highest scores (i.e., scores reflecting worse status), those in aftercare to have lower scores than those in treatment but higher than community members without substance issues. It was felt that community members with substance issues would tend to score more like clients (either those in treatment or in aftercare) than community members with no known substance issues.

4.1.6 Known groups construct validity

A second-order CFA was conducted to test the convergent and discriminant validity of the domains with their convergent indicators. The second-order CFA relates ASA T domain scores and the comparison measure scores (observed variables) to the latent variable or theoretical construct being measured (i.e., daily functioning, relational functioning, dysphoric states, dependence severity, recovery skill/self-efficacy, and existential factors). Of the 242 participants, data were included from all 231 subjects who had complete data on all measures entered into the model. The model yielded a chi-square of 121.3, d.f. = 27, p < 0.001. Examination of the fit indices revealed reasonable fit for both the GFI (0.92) and CFI (0.92), although the RMSEA was disappointing at 0.12 (a model should achieve an RMSEA less than 0.10). This however, may not be unexpected. The RMSEA reflects the model’s covariance structure and is quite sensitive to both high degrees of intercorrelations among latent constructs (which we expected) and sample size. Byrne (2001) notes that when the sample size is small, the RMSEA tends to over-reject true population models. A much greater N might be required to adequately model the high intercorrelations between the constructs reflected in the ASA T.
Table 4

| Sample | M (S.D.) | p value | Effect size: partial η² | Results of post hoc tests
|--------|----------|---------|-------------------------|-----------------------------|
| Treatment, N=142 | Daily functioning | 2.4 (0.70) | <.001 | .35 | T>A= Csub > Cnosub
| Aftercare, N=100 | Relational Functioning | 2.4 (0.68) | <.001 | .23 | T>A= Csub > Cnosub
| Community with substance problems, N=30 | Dysphoric states | 2.6 (0.73) | <.001 | .32 | T>A= Csub > Cnosub
| Community with no apparent substance issues, N=207 | Dependence severity | 2.8 (0.82) | <.001 | .57 | T>A= Csub > Cnosub
| | Recovery skill/self-efficacy | 2.4 (0.69) | <.001 | .29 | T > Csub > A > Cnosub
| | Existential factors | 2.0 (0.68) | <.001 | .62 | T>A> Csub > Cnosub
| | Total ASA T score | 2.4 (0.50) | <.001 | .57 | T>A= Csub > Cnosub

AUDIT: alcohol use disorders identification test; DAST: drug abuse screening test.

Post hoc tests: Tukey's method. All differences are significant at p ≤ 0.001 except for relational functioning, T > A, p = 0.044 and for dysphoric states, Csub > Cnosub, p = 0.049.

The final examination of the ASAT domains was to examine the potential for bias with respect to gender, race (white/minority), and age (younger/older based on a median split). It is possible that a given domain may be detecting true differences that are clinically meaningful. For example, men may in fact have higher mean scores on dependence severity than women, so that this difference reflects actual differences in how addiction manifests in men and women. On the other hand, we do not want the interpretation of ASAT scores to differ for different groups, such that a score on an ASAT domain might suggest problems for one group but not for another. Thus, we compared the ASAT domain scores with each domain's corresponding comparison measure, reasoning that group differences found for an ASAT score, but not for a comparison measure would indicate bias. Bias with respect to a criterion measure requires an examination of slope invariance (Millsap, 1997). Regressions for each domain were run in which the dependent variable was that domain's comparison measure, the subgroup variable (i.e., gender, race, or age), and an interaction term. Bias is assumed if the interaction term is significant in the regression. Regressions were run for each domain for gender, race, and age. No interactions were found to be significant. These results suggest that the ASAT domains do not produce a bias for these subpopulations.

5. Discussion

This article reports on the development of the ASAT, a new measure of problem severity in functional areas relevant to individuals in treatment for addiction. For the field of adult substance abuse treatment, the ASAT provides an outcome measure that is a brief, easy to administer, self-report, multidimensional assessment of addiction related problems that is psychometrically sound and sensitive to change. These qualities make the ASAT ideally suited for outcome studies.
Although not directly tested, the relationship of the ASAT to other measures currently used for treatment planning (i.e., ASI), it is reasonable to speculate that data from the ASAT might be useful for identifying clients’ problem areas that deserve clinical attention. Although such measures exist for mental health and medical functioning, the ASAT offers an outcome assessment tool developed specifically for adult substance abuse clients in treatment. A rigorous developmental process for the ASAT included examination of content validity as determined by substance abuse experts, reliability (internal consistency and test-retest), sensitivity to change, factor structure, convergent and discriminant validity, known groups validity, and bias in terms of gender, race or age.

Clearly, there are limitations of the ASAT. To begin with, the ASAT is a self-report measure. Traditionally, the reliability and validity of self-report from substance abuse clients has been questioned (e.g., Moras, 1993). This concern usually reflects the observed phenomena of “denial” and the consistent under-reporting of consumption in general population surveys. However, for more than a decade, research and reviews continue to support the reliability and validity of self-report of patients entering treatment (e.g., Fuller, 1988; Connors and Maisto, 2003; Del Boca and Darkes, 2003; Neale and Robertson, 2003; Secades-Villa and Fernandez-Hermida, 2003; Yacobian and Urbach, 2002; Solbergsson et al., 2004). Despite generally positive conclusions in the literature regarding the validity of self-report, a few studies have found self-reported use to under-report drug use (e.g., Chermack et al., 2000; Tassiopoulos et al., 2004). Reviewers of the literature seem to agree that: (1) self-report is an important and unique source of information; (2) if assessment conditions are appropriate, self-report of substance abusers tends to be accurate; and (3) self-report should be augmented by information from other sources (Graham et al., 1993).

Second, the ASAT does not constitute a comprehensive assessment. Scores generated by the ASAT are dimensional in nature and are not sufficient to generate DSM/ICD diagnoses or to classify patients according to ASAM criteria. Clarification of the relationship of ASAT data to the results of these other kinds of assessments requires further research. ASAT scores are more relevant to tracking change in patients in treatment or post treatment. The ASAT provides a snapshot of patient status suitable for measuring outcome in important substance abuse-related problem areas, much like the BASIS-32 does for psychiatric populations or the SF-36 provides for medical populations. Another limitation regards the discriminant validity of the ASAT. While the ASAT meets the basic requirements of discriminant validity, the second-order CFA model fit is not robust. Discriminant validity appears to be limited by the degree of intercorrelation among the concepts identified by the substance abuse experts. Consider, for instance, the high degree of intercorrelation among the comparison measures. The average correlation among the comparison measures is 0.40, ranging from 0.20 to 0.57. Indeed, 12 of the 15 pairs among these variables have a correlation greater than 0.30.

Given this level of interrelation among the concepts, it may be mathematically difficult to achieve a high degree of discriminant validity. A trade off may be required between clinical and conceptual relevance of the domains and the demand for psychometric purity. Further research might examine the utility of the conceptual domains of the ASAT for clinical and research purposes. Finally, the sample size in the present study was minimal (Kline, 1999). As discussed, some of the fit indices are sensitive to small sample size (Byrne, 2001) and it is possible that better fit indices would be obtained with a larger sample size.

We also acknowledge that the follow-up rate for the sensitivity-to-change analyses was lower than desirable (68%). It is possible that the data from those lost to follow-up might have reduced the observed effect sizes. However, comparisons of the characteristics of those successfully followed and those lost to follow up did not identify any systematic difference. While it cannot be ruled out that the “improvement” detected by the ASAT is merely regression to the mean, the reasonably good correspondence between change as measured by the ASAT and by the ASI suggests that ASAT-detected change reflect change to a similar degree as another, major outcome assessment currently in use. Furthermore, the baseline and follow-up mean scores obtained in the sensitivity to change study (see Table 2) are virtually identical to the in-treatment and aftercare client groups in the known-groups construct validity comparison (see Table 4). This too provides a fairly strong argument against a regression to the mean explanation. Taken together, these findings support our confidence that the obtained effect sizes are likely meaningful and replicable.

A final point regards the problem acceptance/change readiness domain. While inclusion of an assessment of clients readiness for treatment was supported by the literature and our expert consultants, it became clear that this domain presented certain difficulties for an assessment like the ASAT. Some of this domain’s psychometric parameters were poorer than expected. Perhaps more important, it was difficult to develop items that were not exceedingly dependent upon the respondent’s context. As discussed above, when clients enter treatment, it is expected that they accept the need to make changes in their substance use. Yet, how one answers such questions changes as the respondent moves into early recovery and beyond. As a brief measure focused on problem severity and outcome assessment, the ASAT may not be the best mechanism to measure problem acceptance or change readiness.

The ASAT fills a void in the substance abuse field as a measurement tool that is easy to use for both clients and staff and yields an overview of addiction-related symptoms and functioning concerns of clients in treatment. Researchers and those in clinical settings wishing to conduct outcome studies may find the ASAT useful as a reliable and valid tool that is economical, easy to use, and taps the multidimensional nature of addiction problems. As with comparable measures for psychiatric and medical populations, the ASAT should enable...
users to address the impact of treatment programs on clients' symptoms and functioning status. Given these advantages, we believe the ASAT is a promising new tool for addiction professionals.

Acknowledgments

This project was supported in part by a SBIR Grant DA10131 from the National Institute on Drug Abuse to the first author. Rebecca Cornelli is currently at the Department of Psychology, Loyola University Chicago. The authors would like to thank Traci Craig Green for comments on an earlier draft of this article, Julie Cunningham for assistance with analyses, and Deborah Trottier for her extraordinary data management skills. We also acknowledge the assistance of Drs. Bruce Maslack and Patrice Muchowski. The Addiction Severity Assessment Tool (ASAT) and Manual can be purchased by contacting the first author at Inflexxion, Inc., 320 Needham St., Suite 100, Newton, MA 02464 or via email at sfbutler@inflexxion.com.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version at http://dx.doi.org by entering doi:10.1016/j.drugalcdep.2005.05.005.

References


