



Published in final edited form as:

Eval Rev. 2015 August ; 39(4): 395–427. doi:10.1177/0193841X15599645.

Validation of the Full and Short-Form Self-Help Involvement Scale Against the Rasch Measurement Model

Karen M. Conrad^{1,2}, Kendon J. Conrad^{1,2}, Lora L. Passeti³, Rodney R. Funk³, and Michael L. Dennis³

¹University of Illinois at Chicago School of Public Health, Chicago, IL USA

²Program Metrics, LLC, Oak Park, IL, USA

³Chestnut Health Systems Lighthouse Institute, Normal, IL, USA

Abstract

Background—Substance use disorders (SUDs) are one of the nation’s most costly problems in terms of dollars, disability, and death. Self-help programs are among the varied recovery support options available to address SUD, and evaluation of these programs depends on good measurement. There exists an unmet need for a psychometrically sound, brief, efficient measure of self-help involvement for individuals with SUD that is valid across different substances and age-groups.

Methods—Using data from 2,101 persons presenting for SUD treatment, the full 21-item Global Appraisal of Individual Needs Self-Help Involvement Scale (SHIS) and a newly developed 11-item short-form version were validated against the Rasch measurement model and each other. Differential item functioning (DIF) was assessed by primary substance and age.

Results—Both versions met Rasch psychometric criteria. The full scale had minor misfit with no DIF for alcohol, marijuana, or opioids but a few instances of DIF for amphetamine and cocaine users as well as for age, in that youth tended to endorse several easier items more frequently than did adults. The 11-item short form had neither misfit nor DIF by substance and only minor DIF by age was highly correlated with the full version and was relatively more efficient. Criterion-related validity was supported for both.

Conclusions—Both the long and short versions of SHIS are psychometrically sound measures of a more comprehensive conceptualization of self-help involvement for SUDs that can be used as part of an in-depth assessment or as a short measure that lessens respondent burden.

Keywords

measurement; methodology; substance abuse treatment and prevention

Reprints and permission: sagepub.com/journalsPermissions.nav

Corresponding Author: Karen M. Conrad, University of Illinois at Chicago School of Public Health, 1603 West Taylor Street, Chicago, IL 60612, USA. kmconrad@uic.edu.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The Significance of Self-Help Groups for Substance Use Disorders

Substance use disorders (SUDs) are one of the nation's 10 most costly chronic conditions in terms of dollars, disability, and death (Dennis, Clark, & Huang, 2014). Despite the efficacy of substance use treatment, only a small percentage of those needing treatment access it (Donovan, Ingalsbe, Benbow, & Daley, 2013). However, with the advent of the Affordable Care Act, substance abuse benefits and parity protections now extend to an estimated 62 million Americans (U.S. Department of Health and Human Services, 2013). Data from the National Survey on Drug Use and Health (Substance Abuse and Mental Health Services Administration, 2014) suggest that of the 21.6 million people who had a past year SUD, only 4.1 million received any intervention, including 2.3 million through self-help groups, 1.4 million through outpatient SUD treatment, 1.2 million through outpatient mental health treatment, 1 million through SUD inpatient, and less than a million from any other source. Thus, self-help group involvement has been recognized as an essential component of closing the current gap in SUD treatment and recovery support services (U.S. Executive Office of the President, 2013).

Self-help groups provide a valuable and readily available resource for individuals seeking help with SUDs (Center for Substance Abuse Treatment, 2008; Donovan et al., 2013; Kelly & Yeterian, 2012). They are increasingly recognized as an integral part of managing long-term recovery and as a strategy for lessening the economic burden associated with alcohol and illicit drug use (Bouchery, Harwood, Sacks, Simon, & Brewer, 2011; Dennis, Clark, et al., 2014; Dennis, Foss, & Scott, 2007; Dennis, Scott, & Laudet, 2014; Kelly & Yeterian, 2012; U.S. Department of Justice, 2011). Consequently, it is common for alcohol and drug treatment programs to refer adults and adolescents to these groups both during and after professional treatment (Drug Strategies, 2003; Janchill, 2000; Knudsen, Ducharme, Roman, & Johnson, 2008). Participation in such groups post-treatment is associated with reduced substance use (Kelly, 2013; Kelly, Stout, & Slaymaker, 2013; Pagano, White, Kelly, Stout, & Tonigan, 2013). Further, in our prior work predicting the timing of when people transition from use in the community toward recovery, treatment was one of the biggest predictors; but when we examined who sustained recovery, self-help involvement was the major predictor (Scott, Dennis, & Foss, 2005; Scott, Foss, & Dennis, 2005).

The term self-help groups generally refers to “non-professional, peer-operated organizations devoted to helping individuals who have addiction-related problems” (Humphreys et al., 2004, p. 151). While Alcoholics Anonymous (AA) is the largest and best known self-help organization and uses a 12-step philosophy/approach, a variety of other organizations exist targeting other substances (e.g., Cocaine Anonymous [CA], Narcotics Anonymous [NA]), co-occurring mental health problems (e.g., Double Trouble in Recovery and Dual Recovery Anonymous), family members (Al-Anon and Alateen), and/or with varying philosophies and approaches (e.g., Women for Sobriety [WOS], Secular Organizations for Sobriety [SOS], Self-Management and Recovery Training, among others; Center for Substance Abuse Treatment, 2008; Humphreys et al., 2004).

The Measurement of Self-Help Involvement

While the research evidence documenting the contributions of self-help groups in reducing substance use is encouraging, meta-analyses show inconsistent findings across studies (Ferri, Amato, & Davoli, 2006; Kelly & Yeterian, 2012; Tonigan, Toscova, & Miller, 1996). For example, in their meta-analysis, Tonigan, Toscova, and Miller (1996) noted that “it would be helpful to have greater consensus regarding reliable and valid measurement tools ...” (p. 69). Our ability to draw valid causal inferences about whether a particular self-help program works hinges in part on how well we measure the construct (Shadish, Cook, & Campbell, 2002).

Over the years, researchers have discussed the importance of sound self-help measures. Moos (2008) identified the need to develop reliable and valid measures of the “active ingredients” of substance abuse programs, as did Kingree (1997). More recently, Greenfield and Tonigan (2013) lamented the conceptual and measurement issues that hamper the ability to document program effects.

Researchers have called for more studies that employ measures of self-help group involvement going beyond mere attendance rates (Kelly & Myers, 2007; Kelly, Urbanoski, Hoepfner, & Slaymaker, 2011; Passetti, Godley, & Godley, 2012). A number of Self-Help Involvement Scales (SHIS) meeting this general definition exist. We conducted a literature search of measures that conceived of self-help involvement as a latent construct and that demonstrated acceptable psychometric qualities (e.g., internal consistency reliability and construct validation). Five scales met our criteria all of which were analyzed using classical test theory: the Alcoholics Anonymous Involvement Scale (Tonigan, Connors, & Miller, 1996), the Alcoholics Anonymous Affiliation Scale (Humphreys, Kaskutas, & Weisner, 1998), the Recovery Interview (Morgenstern, Kahler, Frey, & Labouvie, 1996), the General Alcoholics Anonymous Tools of Recovery (Greenfield & Tonigan, 2013), and the 29-item Weekly Self-Help Questionnaire (Weiss et al., 1996). With the exception of Weiss (1996), these measures focus on alcohol abuse and are based on the AA 12-step philosophy. None of these scales were analyzed using item-response theory (IRT)/ Rasch analysis which offers a more powerful examination of item and scale performance than is available using classical test theory alone (Bond & Fox, 2007; Wilson, 2005; Wright & Stone, 1979).

The SHIS, the focus of this article, was developed originally using classical test theory. It is a 21-item scale that taps attendance, participation, and affiliation in self-help groups (Cronbach’s α reported by age-group was .86 to .94). It was explicitly designed for use across types of self-help groups, substances, and ages. The SHIS has been recommended as one of the most reliable, valid, and efficient measures of self-help involvement by the National Institutes of Health Phenx data harmonization project (<https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=560601>).

While the SHIS has been shown to be a reliable and valid measure, further psychometric analysis applying the Rasch measurement model may suggest ways of strengthening and shortening it further. The Rasch model is particularly useful in examining whether items function differently (termed differential item functioning or DIF) across groups, such as

groups categorized by primary substance or by age. If items are found to be invariant, we can be more certain of the validity of score interpretation. The reason for focusing on DIF by primary substance in this article is that prior literature focused on substance specific self-help, for example, for alcohol, narcotics, or CA, while here we are arguing for one measure across substances. The reasons for focusing on DIF by age were that (a) prior analyses have shown substantial DIF by age in both SUD symptoms (K. J. Conrad, Dennis, Bezruczko, Funk, & Riley, 2007) and patterns of use/comorbidity (Chan, Dennis, & Funk, 2008); (b) the rates of accessing self-help are significantly lower among youth than adults (Gangi & Darling, 2012; Kelly & Myers, 1997); and (c) a working paper examining DIF by primary substance, age, gender, and race suggests it was the area of greatest concern (K. M. Conrad, Conrad, Dennis, Riley, & Funk, 2012). To our knowledge, this article is one of the first to evaluate a specific or general measure of self-help involvement using a more rigorous Rasch IRT. Its purpose is to improve the measurement of self-help involvement as a latent, unidimensional construct that is useful across substances and age-groups.

Specific Aims

The specific aims of this study were (1) to evaluate the psychometric properties of the full and a newly developed short-form SHIS using Rasch analysis with a focus on DIF across primary substances and age as well as a traditional test of criterion-related validity and (2) to evaluate the efficiency of the short-form relative to the full SHIS.

The Rasch Measurement Model

A more thorough treatment of the Rasch measurement model is described in detail elsewhere (Bond & Fox, 2007; K. J. Conrad & Smith, 2004). Rasch analysis (Rasch, 1960) is useful in the development of a new scale, revision of an existing scale, or development of a short-form version making it ideal for our purposes. It offers a more powerful examination of item and scale performance than is available using traditional methods, such as classical test theory alone (Bond & Fox, 2007; Wilson, 2005; Wright & Stone, 1979).

Briefly, the Rasch measurement model is the only IRT model that has the desirable scaling properties of linear, interval measurement (Embretson & Reise, 2000). The Rasch model fulfills the requirements of fundamental measurement in that it generates a linear interval scale. Rasch analysis tests whether the data fit the model by assessing whether the response pattern observed in the data corresponds to the theoretical pattern expected by the model (Bond & Fox, 2007). We chose the Rasch model for the following reasons: (1) sound item-fit methods; (2) simplicity and ease of use because only one parameter, that is, difficulty, is needed which is an important consideration for wide clinical use and ease of explanation; (3) scaling properties of linear, interval measurement that enable a clear standard and simple graphic interpretation using person/item maps; (4) eliminates concerns about overfitting with additional parameters and perhaps missing important misfitting patterns, that is, explaining away the misfit with added parameters; and (5) useful for smaller samples because it estimates only a single parameter (K. J. Conrad et al., 2012).

In Rasch analysis, an item hierarchy is created by the item difficulty estimates. Difficulty refers to the relative rarity with which an item is endorsed, with the more rarely endorsed

items considered more difficult (Smith, 2002). In the case of the SHIS, we assessed the logic of the item hierarchy to see if it reflects a progression of involvement from activities that are relatively easy to engage in to those that are harder to accomplish. The term person “ability” refers to the person’s level on the construct being measured, for example, amount of involvement in self-help activities. Unidimensionality is required by the Rasch model and addresses the validity issue of construct-irrelevant variance (Messick, 1995). Item fit statistics are used to evaluate whether individual items contribute to unidimensionality. The model provides for an examination of the data, that is, items and persons, for flaws or problems that are indicated by their failure to fit the model (Bond & Fox, 2007).

As items and persons are placed on the same linear scale, the Rasch model provides a convenient framework for assessing scale invariance across subgroups using DIF analysis (Bond & Fox, 2007; K. J. Conrad et al., 2007). More specifically, DIF analysis enabled us to learn if persons who had similar scores on the SHIS used the item hierarchy differently, for example, arrived at their scores by endorsing items of different difficulty depending on their age-group. A useful example of how DIF can influence scores occurred in examining the Crime and Violence Scale (K. J. Conrad, Riley, Conrad, Chan, & Dennis, 2010). In that study, while some of the males and females received similar scores, they arrived at those scores by endorsing items that were very different in terms of their seriousness. While the females found it easier to endorse the less violent crimes, such as prostitution, theft, and forgery, the males found it easier to endorse armed robbery, assault, and homicide. Despite having similar scores, the males’ scores tended to reflect much more serious crime and violence. Thus, we see that similar scores can have very different meanings.

The SHIS was analyzed using the item-grouping level Masters’ partial credit model (Linacre, 2012). The Rasch model estimates a single parameter, difficulty. In this study, this rating scale model estimates difficulty for each response category within each item grouping. This rating scale model estimates the probability of person n choosing category Level j on Item i of Grouping g as a function of person n ’s position on the latent trait, B_n and the endorsability of Item i of Grouping g at the given threshold, j . In the formula below, the probabilities are converted to log odds, where log is the natural logarithm and the logit is the unit of measure. This allows us to obtain linear, interval-level measures:

$$\log \left[\frac{P_{nij}}{P_{nj(j-1)}} \right] = B_n - D_{gi} - F_{gj},$$

where P is a probability, and the Rasch parameters are B_n , the ability of person, D_{gi} , the difficulty of Item i of Grouping g , and F_{gj} , the Rasch-Andrich threshold between Categories $j - 1$ and j of Grouping g .

Method

Data Source

Respondents completed the SHIS as part of the 3-month posttreatment intake follow-up interview. The SHIS is part of the Global Appraisal of Individual Needs (GAIN; Dennis,

Ives, et al., 2010). The GAIN is a biopsychosocial assessment used in substance abuse treatment and research. All interviewers were trained and certified on using the GAIN using Chestnut's protocol (Titus et al., 2012). This protocol includes standardized feedback on digital recordings of early interviews and ongoing data checks to track the rate of time, interruptions, potential context effects, missing data, inconsistencies between key pairs of items, and evidence of randomness or extreme score answering (based on Rasch model). This was done as part of regular quality assurance regardless of whether the interview was done as part of clinical practice or a research study. Copies of the manual, instruments, psychometrics, norms, current numbers of sites, and list of publications are publicly available at www.gaincc.org. The deidentified data set was downloaded from a secure password-protected website, and institutional review board approval was obtained from Chestnut Health Systems.

The data came from 106 sites across 26 states. The selection of the data was at the individual level: Respondents needed to have completed both the initial GAIN (GAIN-I) and the 3-month follow-up and reported having attended self-help at the 3-month interview.

Sample Characteristics

At the 3-month follow-up interview when the SHIS was administered, most participants were likely receiving treatment or in sober living environments. The 2,103 respondents consisted of a nonrandom sample who ranged in age from 12 to 73 years (mean = 29.2 ± 12.0 ; median = 26; mode = 17). Adults (defined as those at least 18 years of age) comprised 79.6% ($n = 1,675$) of the sample. Males comprised 55.6% ($n = 1,168$) of the sample. Over half of the sample was Caucasian, 56% ($n = 1,177$); 16% ($n = 336$) was Hispanic, 14.7% ($n = 310$) was mixed or other race, and 13.3% ($n = 279$) was African American. Alcohol was the primary drug reported by 21.6% ($n = 453$) of the sample, opioids by 17.6% ($n = 370$); amphetamines by 15.6% ($n = 328$); marijuana by 15.6% ($n = 328$); and cocaine by 10.8% ($n = 226$). For almost a fifth of the sample (18.8%; $n = 396$), the primary drug category was "other." Primary substance was based (in descending order if tied) on the substance with the most recent symptoms, most past year symptoms, most days of use, and participant-identified primary problem. Respondents reported being a regular member or attendee of a variety of self-help groups with the most commonly reported affiliations being AA (52%) or NA (41%). In all, 20% reported no affiliation, 58% reported one affiliation, and 22% reported multiple affiliations. Besides AA and NA, affiliations included CA, Dual Diagnosis Anonymous, LifeRing Secular Recovery, NA, and Secular Organization for Sobriety, Social Recovery, WOS, and Adult Children of Alcoholics.

Measures

Full SHIS—We conceptually define self-help involvement here as the degree of participation or engagement in a self-help group that extends beyond attendance to include helping activities and programmatic activities as well as group affiliation. Consistent with its conceptual definition, the SHIS is operationalized as a 21-item measure including items on the days of self-help meetings attended, behaviors associated with engaging in self-help, and whether the person is "affiliated" with one or more self-help groups.

Of the 21 items within the SHIS, 19 are dichotomous yes/no items that ask respondents about attitudes and behaviors; 1 item asks about days of self-help attendance (coded as 1 = 1–3 days, 2 = 4–24 days, 3 = 25–85 days, and 4 = 86–90 days); and 1 item is a count of self-help group affiliations (coded 0 to 4+ affiliations). The scale stem for the 19 dichotomous items reads “In the past 90 days, have you ... ?” The item stem about days of attendance reads “During the past 90 days, on how many days have you attended one or more self-help group meetings (such as AA, NA, CA, or Social Recovery) for your alcohol or other drug use?” The item about self-help affiliations reads “Do you consider yourself to be a ‘regular attendee or member’ of any specific 12 step fellowships, faith-based, secular or other recovery groups?” The list included those mentioned previously. The item stems, GAIN item numbers, and item labels are displayed in Table 1.

Short-form SHIS—We aimed to develop a short-form SHIS that met the Rasch measurement criteria and at the same time could be used across primary substances and age, was highly correlated with the full scale, was efficient, and could easily be administered and scored. We used an iterative process guided by a strategy used by Hibbard, Mahoney, Stockard, and Tusler (2005). In considering items for deletion, we started with the items that exhibited DIF in the full scale and then we inspected the item hierarchy to be certain that the items sufficiently covered the difficulty range, observed for clusters of items with similar item calibrations, and considered the standard error of measurement for each item. After a series of iterations where we continuously inspected the Rasch output to be sure that the psychometrics were maintained, we deleted 10 items to arrive at the 11-item short form (see Table 1). The details on how we arrived at the short form are presented in the Results section.

Criterion-related measures—Drawing upon the scientific literature that has found an association between self-help group participation and reduced substance use, we selected three criterion measures available in our data set. As noted previously, the SHIS was administered at 3-months posttreatment intake. To establish temporal order, the criterion measures were administered at 6-months posttreatment intake. The recall period for each of the criterion variables is described below along with the hypothesized relationship to the SHIS.

Substance Problem Scale (SPS): The SPS is a 16-item scale (past month) of substance problems operationalized as a count of symptoms of substance abuse, dependence, and substance-induced health and psychological disorders. It is based on the American Psychological Association (APA, 2000) *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV), with 7 items corresponding to DSM-IV criteria for dependence, 4 for abuse, 2 for substance-induced health and psychological problems, and 3 on lower severity symptoms of use (hiding use, people complaining about use, and weekly use). The SPS has demonstrated excellent internal consistency (Cronbach’s α of .90) and good test–retest reliability ($r = .70$; Dennis, Dawud-Noursi, Muck, & McDermeit, 2002; Dennis, Titus, et al., 2002). While the SPS is related to substance use, scores on this scale are conceptually and statistically distinct ($r = .45$ to $.69$; Dennis, Funk, Godley, Godley, &

Waldron, 2004). The Cronbach's α for this sample was .95. We hypothesized a negative correlation with SHIS at $p < .001$.

Substance Frequency Scale (SFS): This is an 8-item scale operationalized as the average percentage of days (in the past 90) in which a respondent reported any alcohol or other drug (AOD) use; heavy AOD use; problems from AOD use; and days of alcohol, marijuana, crack/cocaine, heroin/ opioid, and other drug use. Higher scores represent increasing frequencies of substance use in terms of days, days staying high most of the day (i.e., high risk of problems), and days actually causing problems. The SFS has good internal consistency ($\alpha = .74$ to $.77$) and test-retest reliability ($\rho = .93$; Dennis, Dawud-Noursi, et al., 2002; Dennis, Titus, et al., 2002). The SFS has been shown to be sensitive to the pre-posttest changes associated with outpatient and residential treatment, including randomly controlled trials (Dennis, Dawud-Noursi, et al., 2002; Dennis, Godley et al., 2004). The Cronbach's α for this sample was .83. We hypothesized a negative correlation with SHIS at $p < .01$.

In Recovery: This status variable is scored dichotomously and indicates if a client has been (1) housed in the community, (2) abstinent (any substance) in the past month, and (3) reports no past month substance problems. A value = 1 indicates that the person had no use, abuse, or dependence problems during the past month while living in the community, and is analogous to the DSM-IV concept of early full remission, and a value = 0 indicates that at least one of the three criteria has not been met (Chestnut Health Systems, 2011). Dennis, Scott, and Funk (2003) reported a test-retest $\kappa = .76$. We hypothesized a positive correlation with SHIS at $p < .01$.

Analysis Procedures

We used Winsteps Version 3.75 software program to perform the Rasch analysis (<http://www.winsteps.com/winsteps.htm>) and IBM SPSS Statistics Release 21 to examine criterion validity. A summary of how we analyzed each criterion is discussed subsequently.

Aim 1: Evaluate the Psychometric Properties of the SHIS

Assess unidimensionality—Principal component analysis of residuals was used to examine whether a substantial factor existed in the residuals after the primary measurement dimension has been estimated (Smith, 2002). This allowed us to determine whether the set of items represented a single construct as intended. Analysis of the dimensionality involved a two-step process. First, the measurement dimension of the scale was estimated using the Rasch model. The variance associated with this measurement dimension was extracted from the item-response data by computing standardized residuals, that is (observed – expected)/ (model standard error). Second, a principal component analysis of the standardized residuals was used to determine whether substantial subdimensions existed within the items (Smith, 2002). If the items measure a single latent dimension as estimated by the Rasch model, then the remaining residual variance should reflect random variation (Smith, 2002).

Our first criterion for unidimensionality was that the variance explained by the measurement dimension be large, that is, at least 40% (Linacre, 2006; Reckase, 1979). The second

criterion was that the variance explained by the first principal component of the residuals be small (i.e., no more than 15%). We also chose a minimum ratio of 3:1 for the variance in the measurement dimension compared to the variance of the first principal component of residuals (Embretson & Reise, 2000). We employed additional criteria for unidimensionality using item-fit statistics, discussed subsequently.

Examine response category functioning—When using the rating scale model with items having more than two response options, such as a Likert-type format, it is important to examine the response category structure to evaluate how the response options are used (Bond & Fox, 2007). If the thresholds do not progress in a linear fashion, or the distances between two response category thresholds are judged to be inadequate, the category functioning may be improved by collapsing response categories (Bond & Fox, 2007).

Assess item fit—Rasch analysis provides fit statistics to test assumptions of fundamental measurement (Bond & Fox, 2007), for example, that high scorers should endorse or get right almost all of the easy items. Once identified, items that “misfit” can then be examined qualitatively to determine the causes of the problems, for example, confusing wording or items that assess a construct that is different from the principal one being measured. Understanding poor fit can guide decisions about improving or dropping items. The Rasch model provides two indicators of misfit: infit and outfit, where infit refers to inlying improbable or random patterns and outfit to highly unexpected outlier responses. Mean square (MNSQ) fit statistics are defined such that the model-specified uniform value of randomness is 1.0 (Bond & Fox, 2007), indicating good fit to the Rasch model. In item-fit analysis, MNSQ values between .75 and 1.33 are considered acceptable (Wilson, 2005). We considered that an item misfit if both infit MNSQ and outfit MNSQ > 1.33.

Assess person reliability—The Rasch person reliability is based on the estimated locations of persons along the measurement continuum. Thus, while similar to Cronbach’s α , person reliability is often lower because, unlike Cronbach’s α , it does not include extreme scores (i.e., zero or perfect scores) in its computation. The values are interpreted in the same manner. For both indices, our criterion for acceptability in an existing measure was .80 and for a short measure was .75.

Examine the item hierarchy—In Rasch analysis, the item hierarchy that is created by the item difficulty estimates provides an indication of construct validity, that is, the items are ordered in terms of difficulty or endorsability (K. J. Conrad et al., 2012). Rasch analysis allows persons and items, measured on the same logit scale, to be displayed simultaneously on a Wright Item Map. In developing the SHIS, a theory was not articulated to guide an item hierarchy. The notion of hypothesizing an a priori item hierarchy is recommended when developing scales using IRT/Rasch (Wilson, 2005); however, it is not in classical test theory. The Rasch analysis in the current study provided an opportunity to examine the item hierarchy.

Assess DIF for primary drug and for age—The Rasch model requires that subgroups having equal levels of the underlying construct, that is, same or similar score, should respond in a similar manner to the items measuring this construct (Bond & Fox, 2007). In

this case, examination of DIF allowed us to test whether items reflect significantly different levels of self-help involvement for different groups, for example, different item calibrations for youth versus adults. For example, in the case of age DIF, we wanted to learn if youth and adults with the same or similar scores, rated a given item as being of similar difficulty to endorse, easier to endorse, or harder to endorse, for example, Did youth and adults find the item *performing services at a meeting* equally difficult to endorse? Items that demonstrate DIF should be investigated to determine what may be inferred about the underlying construct and what that implies about the samples of persons (Bond & Fox, 2007). We used a criterion of $DIF > 0.6$ logit difference in judging a meaningful DIF contrast. Standards for what is considered an important DIF effect size vary from about .4 to .6 logits (Draba, 1977; Elder, McNamara, & Congdon, 2003; Longford, Holland, & Thayer, 1993; Paek, 2002; Scheuneman & Subhiyah, 1998; Wang, 2000). If the items are not invariant across subgroups, then changes to the measures can be considered, for example, dropping items, developing new ones, or developing separate measures for specific subgroups.

Examine criterion-related construct validity—We examined predictive validity (Messick, 1995) using theory-based directional hypotheses (i.e., Pearson one-tailed bivariate correlations) between the SHIS, measured at 3 months posttreatment intake, and the criterion variables measured at 6 months post-treatment intake; p values less than .01 were considered statistically significant. In addition, correlations greater than .5 were considered large; .5 to .3 were considered moderate; .3 to .1 were considered small; and anything less than .1 was considered trivial (Cohen, 1988).

Aim 2: Compare the Efficiency of the Short-Form Relative to the Full SHIS

Our measure of efficiency was defined as follows:

$$\text{Efficiency} = (\# \text{ of short-form items} / \# \text{ of full scale items}) / (\text{correlation between the full and short form scales.})$$

The goal is to create a short-form scale with the minimum number of items while maintaining a high correlation between the full and short-form versions. Thus, in the formula, the smaller the numerator the better, and the greater the denominator the better, making for a more efficient short-form scale (Dennis, Chan, & Funk, 2006). Our criterion for efficiency was .80.

Results

Aim 1a: Evaluate the Psychometric Properties of the Full SHIS

Assess unidimensionality—In assessing unidimensionality of the original 21-item SHIS, the variance explained by the measure was 50.3%, and the variance explained by the first principal component of residuals was 11.9%, forming a ratio of about 4:1. These results met our criteria for unidimensionality. Examination of dimensionality revealed one clear dimension that made sense theoretically.

Examine response category functioning—Next, we examined the category, aka, step functioning for the 2 items with more than one step. Put simply, each response category

should rise monotonically or step up, and the size of the rise should be reasonably large in order to be efficient, that is, 1.4 to 5 logits (Linacre, 1999). This is not pertinent to the 19 dichotomous items (Items #2–#20) because they only have one step and all were functioning in the expected direction, that is, no reversed items. Item #1 “on days attended” had a 1–4 response format (3 steps) and formed the large, monotonic progressions from one step to the next. Item #21 “number of affiliations” had a 0–4 response format (4 steps). However, the response category function for this item showed a confused pattern. While we observed the desired monotonic progression from one step calibration to the next, the interval between step calibrations for the 3 and 4 response options was much smaller than the recommended 1.4 to 5 logit interval (Linacre, 1999). Only 4% of respondents selected Response options 3 or 4. We therefore combined Response options 2, 3, and 4 resulting in a recoded 3-point response format, that is, 0, 1, and 2. The intervals between step calibrations for the response options subsequently met the recommended 1.4 to 5 logit interval (Linacre, 1999) conforming to the recommended Rasch Andrich thresholds (Andrich, 1978).

Assess item fit—We proceeded with the analysis using the recoded #21 on number of affiliations. Two items misfit based on our criterion of >1.33 for both infit and outfit: Item #1 (days attended) and Item #21 (number of affiliations). As noted earlier, both of these items had a response format different from the other 19 dichotomous items. At this time in the analysis, we decided to retain these items given that the misfit may have been due to a method artifact and our judgment that including items on days and affiliations was important to the scale.

Assess person reliability—The Rasch person reliability was .88 and Cronbach’s α was .90.

Examine item hierarchy—The Wright map for the full 21-item scale is depicted in Figure 1 (Map A). The left-hand side of the map shows the distribution of respondents in our sample arranged along a hierarchy by person ability, that is, their level of involvement in self-help. The right-hand side of the map shows the array of items ordered by item endorsability, that is, from “easy to endorse” to “hard to endorse.” The Item Map shows that the items were well targeted to the sample. Persons at the top have the highest level of the construct being measured, and items at the top of the map are the most difficult to endorse (Bond & Fox, 2007). The item hierarchy was very stable as indicated by item reliability of 1.0 (Wright & Stone, 1979). The hierarchy supported our notion about the progression of involvement from lower levels of involvement, such as Item 11, You Understood Others and Item 10, Others Understood You to activities that reflect a higher degree of involvement, such as Item 16, Sponsor Someone.

Assess DIF—We examined DIF for the five primary substances of alcohol, amphetamines, cocaine, marijuana, and opioids. Remaining substances were classified as “other.” In conducting the DIF analysis, dichotomous comparisons were conducted such that each of the five primary substances was compared to all of the other primary drugs combined. Using our criterion of >0.6 logit difference, there were no instances of DIF for alcohol, marijuana, or opioids. There were four instances of DIF for amphetamine and

cocaine (DIF contrast range .78 to .92). Amphetamine users were less likely to endorse Item 13, Agreed Advice and Item 16, Sponsor Someone. Cocaine users were less likely to endorse Item 7, Reading Recovery but more likely to endorse Item 9, Pray Help. We did not view these few instances of DIF as causing serious distortion of measurement. This was especially true for cocaine since the DIF balanced for the 2 items, one easier and one harder.

Age was classified as youth (younger than 18 years) or adult (18 years and older). Four items exhibited DIF (DIF contrast range .63 to .78). Youth were more likely to endorse Item 7, Reading Recovery; Item 11, You Understood Others; Item 13, Agreed Advice; and Item 17, Perform Service Meeting. Because all the items that met the criterion for DIF were easier for youth to endorse, youth would tend to get their scores by endorsing easier items.

Examine criterion-related construct validity—As hypothesized, we observed negative correlations between the 21-item SHIS and the Substance Problem Scale ($r = -.10$) and the Substance Frequency Scale ($r = -.12$). As hypothesized, we observed a statistically significant positive correlation with the In Recovery status measure ($r = .24$). Using Cohen's (1988) criteria, we classified these correlations as small. All correlations were statistically significant at $p < .000$.

Summary for the Full SHIS

Rasch analysis results suggested that the full 21-item SHIS performed well as a unidimensional measure of self-help involvement with items well targeted to the sample. Internal consistency was good with a Rasch person reliability of .88 and a Cronbach's α of .90. The minor misfit in 2 items was likely due to a method artifact. Of the 21 items in the SHIS, meaningful DIF occurred in 4 items for amphetamines and cocaine and in 4 items by age. We concluded that the presence of DIF was probably not sufficient to undermine the general validity of the measure. However, the DIF results were useful in suggesting items for deletion for the short-form version. Construct validity was demonstrated within the limitations acknowledged subsequently.

Aim 1b: Evaluate the Psychometric Properties of the Short-Form SHIS

In arriving at the short form, first, we deleted the 6 items that exhibited DIF in the full 21-item version: Item 7, Reading Recovery; Item 9, Pray Help; Item 11, You Understood Others; Item 13, Agreed Advice; Item 16, Sponsor Someone; and Item 17, Perform Service Meeting. Item 16, Sponsor Someone, also had the highest standard error (.14). Next, we deleted Item 21, Number SH Affiliations, which had the highest misfit and required a scoring algorithm. Finally, 3 items were deleted because they occupied essentially the same location on the item difficulty hierarchy as another item(s), thus providing minimal additional information: Item 4, Talk Sponsor at Meeting; Item 5, Talk Sponsor/Other Out; and Item 15, Help Someone Meeting.

Assess unidimensionality—The variance explained by the measure was 49.8%, and the variance explained by the first principal component of residuals was 14.7%, forming a ratio of 3.4:1. These results met our criteria for unidimensionality.

Assess response category functioning—The items formed the desired monotonic progression from one step to the next, and the response curves formed discrete peaks for each response option.

Assess item fit—No items misfit based on our criterion of >1.33 for both infit and outfit.

Assess person reliability—The Rasch person reliability was .77 and Cronbach's α was .83.

Examine item hierarchy—The Wright map for the 11-item scale is depicted in Figure 1 (Map B). With only a slight floor effect, the sample was well targeted to the items indicating good targeting for this SUD population.

Assess DIF for primary drug and age—There was no evidence of DIF by primary substance. There was minor DIF for age where age was classified as youth (younger than 18 years) or adult (18 years and older). Youth were more likely to endorse Item 12, Received Advice (DIF contrast .62) and Item 18, Participate SH Events (DIF contrast .66). Since these 2 items did not have significant DIF on the 21-item version and since their DIF contrasts were very near our criterion of $>.6$ for DIF, we regarded these as possible chance findings that should be examined in future studies of other samples. Therefore, the removal of the 4 DIF items from the 21-item scale that were easier for youth to endorse provided substantial improvement in reducing or removing the potential measurement distortion for youth compared to adults for the short form.

Examine criterion-related construct validity—Hypotheses for the relationship between the 11-item SHIS and the three criterion-related variables were supported. As hypothesized, we found statistically significant negative correlations between the 11-item SHIS and the Substance Problem Scale ($r = -.11$) and the Substance Frequency Scale ($r = -.13$). As hypothesized, we observed a statistically significant positive correlation with the Recovery Index ($r = .23$). Using Cohen's (1988) criteria, we classified these correlations as small. All correlations were statistically significant at $p < .001$.

Summary for the Short-Form SHIS

The 11-item short form met the Rasch criteria being a valid unidimensional measure of self-help involvement. Although a slight floor effect was present, the items were generally well targeted to the sample. Internal consistency was good with a Rasch person reliability of .77 and a Cronbach's α of .83. The scale had neither misfit nor DIF by substance but minor DIF by age.

Aim 2: Evaluate the Efficiency of the Short-Form Relative to the Full SHIS

In comparing the psychometrics of the short-form to the full SHIS, we observed a slight decrease in internal consistency, that is, Rasch person reliability dropped to .77 from .88 and Cronbach's α dropped to .83 from .90. Although both versions met our criteria for unidimensionality, the values for the short form were less strong. After deleting the items that exhibited DIF in the full version, the short form showed minor DIF by age only. The

correlations with the criterion measures were essentially unchanged all being statistically significant but small.

With both versions meeting our psychometric criteria, we proceeded to evaluate the efficiency of the short form. The full 21-item and 11-item short form were highly correlated ($r = .97$). With 10 items deleted, the short form consisted of 52% of the original 21 items. Using the efficiency formula described earlier, $.52/.97$ produces an efficiency score of $.54$, thus meeting our criterion of $.80$. Therefore, the short form was relatively more efficient. Of course, fewer items are more efficient only to the extent that critical information is not lost.

Discussion

Rasch analysis supported the psychometric properties of both the full 21-item and short 11-item SHIS finding both versions of the scale to be unidimensional, with good person reliability, and valid. Minor misfit in 2 items (days of attendance and number of affiliations) was likely due to a method artifact because they were on a different response scale than the other items (which were all yes/no). Because of the high correlation between the long and short version of the scales, the latter was much more efficient. If the SHIS is being used as one of many general measures (vs. the focus of the study), end users may want to use the shorter version. But it should be noted that if the focus of a study is on self-help or it is being done in a new setting or population, it may still be worthwhile to use the longer version to take advantage of the slightly higher reliability and enable more detailed examination of particular self-help involvement aspects (e.g., talking to sponsor, reading recovery-related materials, or performing services at meeting).

Item Hierarchy

The general ordering of the SHIS items from those that were easier to endorse to those that were harder to endorse makes clinical sense. We found that *understanding others* and having *others understand you* were easier for individuals in the overall sample to endorse before endorsement of self-help programmatic services, such as talking to a sponsor, working the steps, and becoming a member of a home group. It also makes sense that *sponsoring someone* was the hardest item to endorse since that usually occurs after an extended period of abstinence and group involvement.

This item ordering is consistent with that of other researchers who examined the frequency of engagement in selected self-help activities. In separate studies of adults and adolescents, Pagano, Krentzman, et al. (2010) and Pagano, Kelly, et al. (2013) found in the validation of an AA-related helping scale, Service to Others in Sobriety, that the most frequently reported activities were fellowship-oriented helping activities, such as listening to an addict or putting away chairs, while the least frequently reported were programmatic services, such as held a service position. Kelly, Myers, and Rodolico (2008) also concluded that general group therapeutic factors, rather than 12 step-specific ones, were most valued by adolescents during the early stages of recovery.

DIF

Though we are unclear as to why, we did find some DIF when the primary substance was a stimulant (amphetamine or cocaine) compared with those who reported other substances despite having a similar scale score. More specifically, amphetamine users were less likely to *agree with advice* and *sponsor someone*. Cocaine users were less likely to *read recovery materials* but more likely to *pray for help*, which balances out any potential measurement distortion. Each substance was compared to all other substances combined, complicating interpretation. These findings may merit further investigation, but we regard them as minor based on these data. In reality, the small amount of DIF in the full version likely does not compromise the scale's validity and use across primary substances.

Regarding age DIF, we noted that there would be a tendency for youth to endorse easier items more frequently than adults. This would make youth tend to appear higher in involvement than adults with similar scores. Adults, however, were endorsing the harder items more frequently which reflected a higher level of involvement. Specifically, compared to adults with the same/similar score, youth were more likely to endorse 3 easy items *read recovery materials* (item measure = -1.38), *understand others* (item measure = -3.16), *agree with advice* (item measure = -1.63), and one harder item, *perform services at meeting* (item measure = 1.65). Because these 4 items were deleted in the short form, the short form should correct substantially for the potential measurement distortion by age.

While we were not able to examine the mix of adults/adolescents in the different affiliation groups, Kelly and Myers (1997) found that 65% of the 12-step meetings attended by adolescents in their study were comprised mainly of older individuals. Adults often differ from youth in life experiences and substance-use patterns and related problems (Stewart & Brown, 1995) and tend to have less substance-abuse problem recognition and motivation for abstinence (Tims et al., 2002). On the other hand, clinicians have reported that some youths seem to prefer adult meetings for the wisdom and praise provided by older individuals (Passetti & Godley, 2008).

It was surprising that adolescents were more likely to endorse *agreeing with advice* given in meetings because of a common clinical concern that youths may have difficulty understanding and embracing self-help group concepts framed in adult language in recovery support groups, particularly ones grounded in the 12 steps (Deas & Thomas, 2001; Passetti & Godley, 2008). Twelve-step concepts like "acceptance," "surrender," and "spirituality" have been identified as potentially too abstract for many adolescents to grasp (Deas & Thomas, 2001; Passetti & Godley, 2008).

We conclude that youth will tend to have elevated scores due to endorsement of easier items. Adults will tend to endorse harder, service-type items on the long form. This may not be problematic and may even be desirable if a youth-sensitive measure is desired. In any case, we should recognize that youth scores will tend to represent a slightly different type of involvement than do adult scores. If we want scores that are comparable for youth and adults, the short form will provide that comparability.

Future Research

Further measurement work is needed such as DIF analysis across more age categories and self-help affiliation groups. Regarding age-groups, Chi, Campbell, Sterling, and Weisner (2012) recently called for additional research that considers how the relationship between self-help and beneficial outcomes may vary by adolescent subgroups. In addressing this research gap, when sample size permits, it may be useful to further subdivide the age-groups based on developmental milestones when exploring DIF contrasts to determine how these subgroups use the measure. Using Rasch to conduct a DIF analysis of alcohol involvement by age, Kahler, Hoepfner, and Jackson (2009) found a drift in the relative ordering of items from adolescence to young adulthood that was consistent with developmental changes.

It would be useful to understand if the SHIS can predict substance use outcomes. A significant percentage of individuals who join self-help groups drop out (Kelly & Moos, 2003), and it may help clinicians better facilitate self-help involvement if the SHIS can also predict who will drop out. Further evaluations of treatments that include self-help components could use SHIS to shed light on whether or not scale scores and patterns of item endorsement change during various stages of recovery and relapse, levels of care, and severities of substance use.

Limitations

The SHIS is a self-report measure with all the limitations of self-report, such as the possibility of socially desirable responses. Cross-validation of responses with clinicians, sponsors, and other collateral informants was not available. The psychometric analysis was conducted using only the 3-month postbaseline SHIS measure when most participants were likely to be receiving treatment or in sober living environments. The short-form SHIS was derived using the same sample that completed the full SHIS, rather than an independent sample. We also note that along with the SHIS, the measures used for criterion validation were part of the GAIN, thus shared method variance may have tended to increase correlations. Further criterion validation with external measures under a variety of methods and settings should be conducted.

The sample size was large and diverse in terms of age, race, age, and primary substance use, but it was restricted to persons entering SUD treatment. Although the GAIN data set is one of the largest samples available to date, it is not a random sample of all people entering SUD treatment and thus may not be representative nor does it reflect people who attend self-help groups without also receiving formal treatment services. While here we focused on primary substance and age and in earlier analyses we examined gender and race (K. M. Conrad et al., 2012), it is possible that other variables (e.g., referral source, comorbidity) might also matter. Finally, it was not possible to isolate the role of specific self-help affiliation groups, given that more than three quarters of respondents reported none or multiple affiliations.

Conclusion

While recommendations and requirements to attend self-help groups are already popular among treatment programs for substance use problems, we can expect an increase in such

referrals by a variety of health care professionals (Donovan et al., 2013) due to recent enacted legislation. The finding of DIF by primary substance and age illustrated the importance and usefulness of developing a common metric for evaluating self-help involvement. This is significant because in practice most people presenting to both treatment and self-help groups use multiple substances, over three quarters of the people attended multiple types of groups and most people with SUD use from adolescents into adulthood.

Both the full and short-form SHIS versions met Rasch psychometric criteria. Having these psychometrically sound, efficient self-help measures available for use across primary substances and age-groups can help us advance our understanding of how self-help groups may work to reduce substance abuse. This article has demonstrated that both the long and short versions of SHIS are psychometrically sound measures of a more comprehensive conceptualization of self-help involvement that can be used as part of an in-depth assessment or as a short measure to lessen respondent burden.

Acknowledgments

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The proposed work was supported by NIDA grant no. R37-DA011323 (PI: M. Dennis), NIAAA grant no. R01 AA021118 (PI: M. Godley), and SAMHSA Contract #270-12-0397 (PD: M. Dennis).

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000.
- Andrich D. Rating formulation for ordered response categories. *Psychometrika*. 1978; 43:561–573.
- Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. 2. Mahwah, NJ: Lawrence Erlbaum; 2007.
- Bouchery EE, Harwood HJ, Sacks JJ, Simon CJ, Brewer RD. Economic costs of excessive alcohol consumption in the U.S., 2006. *American Journal of Preventive Medicine*. 2011; 41:516–524. <http://dx.doi.org/10.1016/j.amepre.2011.06.045>. [PubMed: 22011424]
- Center for Substance Abuse Treatment. An introduction to mutual support groups for alcohol and drug abuse: Substance abuse in brief fact sheet. Rockville, MD: 2008. DHHS Publication Number SMA 08-4336 Retrieved from <http://www.samhsa.gov/shin>
- Chan YF, Dennis ML, Funk RR. Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment. *Journal of Substance Abuse Treatment*. 2008; 34:14–24. [PubMed: 17574804]
- Chestnut Health Systems. GAIN-I encyclopedia of supplemental documentation on scales and other calculated variables [Electronic version]. Normal, IL: Chestnut Health Systems GAIN Coordinating Center; 2011. Retrieved from http://www.gaincc.org/_data/files/Psychometrics_and_Publications/Resources/GAIN-I_Scales_and_Variables.xls
- Chi FW, Campbell CI, Sterling S, Weisner C. Twelve-step attendance trajectories over 7 years among adolescents entering substance use treatment in an integrated health plan. *Addiction*. 2012; 107:933–942.10.1111/j.1360-0443.2011.03758.x [PubMed: 22151625]
- Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: L. Erlbaum Associates; 1988.
- Conrad KJ, Conrad KM, Mazza J, Riley BB, Funk R, Stein MA, Dennis ML. Dimensionality, hierarchical structure, age generalizability, and criterion validity of the GAIN's behavioral complexity scale. *Psychological Assessment*. 2012; 24:913–924.10.1037/a0028196 [PubMed: 22545694]

- Conrad KJ, Dennis ML, Bezruczko N, Funk RR, Riley BB. Substance use disorder symptoms: Evidence of differential item functioning by age. *Journal of Applied Measurement*. 2007; 8:373–387. [PubMed: 18250524]
- Conrad KJ, Riley BB, Conrad KM, Chan YF, Dennis ML. Validation of the Crime and Violence Scale (CVS) against the Rasch measurement model including differences by gender, race, and age. *Evaluation review*. 2010; 34:83–115.10.1177/0193841X10362162 [PubMed: 20233998]
- Conrad KJ, Smith EV Jr. International conference on objective measurement: Applications of Rasch analysis in health care. *Medical care*. 2004; 42:11–16.10.1097/01.mlr.0000103527.52821.1c [PubMed: 14707750]
- Conrad, KM.; Conrad, KJ.; Dennis, ML.; Riley, BB.; Funk, R. Validation of the Self Help Involvement Scale (SHIS) to the Rasch Measurement Model (GAIN Methods Report 1.0). Chicago, IL: Chestnut Health Systems; 2012.
- Deas D, Thomas SE. An overview of controlled studies of adolescent substance abuse treatment. *American Journal on Addictions*. 2001; 10:178–189. [PubMed: 11444159]
- Dennis ML, Chan YF, Funk RR. Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *The American Journal on Addictions*. 2006; 15:80–91.10.1080/10550490601006055 [PubMed: 17182423]
- Dennis ML, Clark HW, Huang LN. The need and opportunity to expand substance use disorder treatment in school-based settings. *Advances in School Mental Health Promotion*. 2014; 7:75–87.10.1080/1754730x.2014.888221
- Dennis, ML.; Dawud-Noursi, S.; Muck, R.; McDermeit, M. The need for developing and evaluating adolescent treatment models. In: Stevens, SJ.; Morral, AR., editors. Adolescent substance abuse treatment in the United States: Exemplary models from a national evaluation study. Binghamton, NY: Haworth Press; 2002. p. 3-34.
- Dennis ML, Foss MA, Scott CK. An eight-year perspective on the relationship between the duration of abstinence and other aspects of recovery. *Evaluation Review*. 2007; 31:585–612.10.1177/0193841X07307771 [PubMed: 17986709]
- Dennis ML, Funk R, Godley SH, Godley MD, Waldron H. Cross-validation of the alcohol and cannabis use measures in the Global Appraisal of Individual Needs (GAIN) and Timeline Followback (TLFB; Form 90) among adolescents in substance abuse treatment. *Addiction*. 2004; 99:120–128.10.1111/j.1360-0443.2004.00859.x [PubMed: 15488110]
- Dennis ML, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J, Funk R. The Cannabis Youth Treatment (CYT) Study: Main findings from two randomized trials. *Journal of Substance Abuse Treatment*. 2004; 27:197–213. S0740-5472(04)00087-X [pii]. 10.1016/j.jsat.2003.09.005 [PubMed: 15501373]
- Dennis, ML.; Ives, M.; Funk, R.; Modisette, K.; Bledsaw, R.; Innes, P. GAIN-I encyclopedia of supplemental documentation on scales and other calculated variables. 2010. Retrieved from http://www.chestnut.org/LI/gain/Scales_Crosswalks/GI_Scales_and_Variables.xls
- Dennis ML, Scott CK, Funk R. An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders. *Evaluation and Program Planning*. 2003; 26:339–352. [http://dx.doi.org/10.1016/S0149-7189\(03\)00037-5](http://dx.doi.org/10.1016/S0149-7189(03)00037-5).
- Dennis ML, Scott CK, Laudet A. Beyond bricks and mortar: Recent research on substance use disorder recovery management. *Current Psychiatry Reports*. 2014; 16:442. [PubMed: 24557873]
- Dennis ML, Titus JC, Diamond G, Donaldson J, Godley SH, Tims FM, The CYTSC. The Cannabis Youth Treatment (CYT) experiment: Rationale, study design and analysis plans. *Addiction*. 2002; 97:16–34.10.1046/j.1360-0443.97.s01.2.x [PubMed: 12460126]
- Donovan DM, Ingalsbe MH, Benbow J, Daley DC. 12-step interventions and mutual support programs for substance use disorders: An overview. *Social Work in Public Health*. 2013; 28:313–332.10.1080/19371918.2013.774663 [PubMed: 23731422]
- Draba, RE. The identification and interpretation of item bias (Research Memorandum No. 26). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago; 1977.
- Drug Strategies. Treating teens: A guide to adolescent drug programs. Washington DC: Author; 2003.

- Elder C, McNamara T, Congdon P. Rasch techniques for detecting bias in performance assessments: An example comparing the performance of native and non-native speakers on a test of academic English. *Journal of Applied Measurement*. 2003; 4:181–197. [PubMed: 12748409]
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
- Ferri, M.; Amato, L.; Davoli, M. *Cochrane Database of Systematic Reviews*. 2006. Alcohol anonymous and other 12-step programmes for alcohol dependence; p. Art No. CD005032
- Gangi J, Darling CA. Adolescent substance-use frequency following self-help group attendance and outpatient substance abuse treatment. *Journal of Child & Adolescent Substance Abuse*. 2012; 21:293–309. <http://dx.doi.org/10.1080/1067828X.2012.702937>.
- Greenfield BL, Tonigan JS. The general alcoholics anonymous tools of recovery: The adoption of 12-step practices and beliefs. *Psychology of Addictive Behaviors*. 2013; 27:553–561. <http://dx.doi.org/10.1037/a0029268>. [PubMed: 22867293]
- Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. [Research Support, Non-U.S. Gov't]. *Health Services Research*. 2005; 40:1918–1930. [PubMed: 16336556]
- Humphreys K, Kaskutas LA, Weisner C. The alcoholics anonymous affiliation scale: Development, reliability, and norms for diverse treated and untreated populations. *Alcoholism: Clinical and Experimental Research*. 1998; 22:974–978. 10.1111/j.1530-0277.1998.tb03691.x
- Humphreys K, Wing S, McCarty D, Chappel J, Gallant L, Haberle B, Weiss R. Self-help organizations for alcohol and drug problems: Toward evidence-based practice and policy. *Journal of Substance Abuse Treatment*. 2004; 26:151–158. 10.1016/s0740-5472(03)00212-5 [PubMed: 15063905]
- Janchill N. Substance dependency treatment for adolescents: Practice and research. *Substance Use and Misuse*. 2000; 35:2031–2060. [PubMed: 11138716]
- Kahler CW, Hoepfner BB, Jackson KM. A Rasch model analysis of alcohol consumption and problems across adolescence and young adulthood. *Alcoholism: Clinical & Experimental Research*. 2009; 33:663–673.
- Kelly JF. Alcoholics anonymous science update: Introduction to the special issue. [Introductory]. *Substance Abuse*. 2013; 34:1–3. [PubMed: 23327498]
- Kelly JF, Moos RH. Dropout from 12-step self-help groups: Prevalence, predictors, and counteracting treatment influences. *Journal of Substance Abuse Treatment*. 2003; 24:241–250. 10.1016/s0740-5472(03)00021-7 [PubMed: 12810145]
- Kelly JF, Myers MG. Adolescent treatment outcome in relation to 12-step group attendance. *Alcoholism: Clinical and Experimental Research*. 1997; 21:27A.
- Kelly JF, Myers MG. Adolescents' participation in alcoholics anonymous and narcotics anonymous: Review, implications and future directions. *Journal of Psychoactive Drugs*. 2007; 39:259–269. [PubMed: 18159779]
- Kelly JF, Myers MG, Rodolico J. What do adolescents exposed to alcoholics anonymous think about 12-step groups? *Substance Abuse*. 2008; 29:53–62. [PubMed: 19042324]
- Kelly JF, Stout RL, Slaymaker V. Emerging adults' treatment outcomes in relation to 12-step mutual-help attendance and active involvement. [Research Support, N.I.H., Extramural]. *Drug & Alcohol Dependence*. 2013; 129:151–157. [PubMed: 23122600]
- Kelly JF, Urbanoski K, Hoepfner B, Slaymaker V. Facilitating comprehensive assessment of 12-step experiences: A multidimensional measure of mutual-help activity. *Alcoholism Treatment Quarterly*. 2011; 29:181–203. <http://dx.doi.org/10.1080/07347324.2011.586280>. [PubMed: 22081741]
- Kelly JF, Yeterian JD. Empirical awakening: The new science on mutual help and implications for cost containment under health care reform. *Substance Abuse*. 2012; 33:85–91. 10.1080/08897077.2011.634965 [PubMed: 22489579]
- Kingree JB. Measuring affiliation with 12-step groups. *Substance Use and Misuse*. 1997; 32:181–194. [PubMed: 9044546]
- Knudsen, HK.; Ducharme, LJ.; Roman, PM.; Johnson, JA. *Service delivery and use of evidence-based treatment practices in adolescent substance abuse treatment settings: Project report*. Rober Wood

- Johnson Foundation's Substance Abuse Policy Research Program Grant No. 53130. 2008. Retrieved from <http://ntcs.uga.edu/reports/Adolescent%20Study%20Summary%20Report>
- Linacre JM. Investigating rating scale utility. *Journal of Outcome Measurement*. 1999; 3:103–122. [PubMed: 10204322]
- Linacre JM. Data variance explained by measures. *Rasch Measurement Transactions*. 2006; 20:1045–1047.
- Linacre, JM. Winsteps Rasch measurement (Version 3.75). 2012. Retrieved from <http://winsteps.com>
- Longford, NT.; Holland, PW.; Thayer, DT. Stability of the MH D-DIF statistics across populations. In: Wainer, PWHH., editor. *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 67-113.
- Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50:741–749.
- Moos RH. Active ingredients of substance use-focused self-help groups. *Addiction*. 2008; 103:387–396. [10.1111/j.1360-0443.2007.02111.x](https://doi.org/10.1111/j.1360-0443.2007.02111.x) [PubMed: 18269361]
- Morgenstern J, Kahler CW, Frey RM, Labouvie E. Modeling therapeutic response to 12-step treatment: Optimal responders, non-responders, and partial responders. *Journal of Substance Abuse*. 1996; 8:45–59. [PubMed: 8743768]
- Paek, I. Unpublished doctoral dissertation. Berkeley: University of California; 2002. Investigations of differential item functioning: Comparisons among approaches, and extension to a multidimensional context.
- Pagano ME, Kelly JF, Scur MD, Ionescu RA, Stout RL, Post SG. Assessing youth participation in AA-related helping: Validity of the service to others in sobriety (SOS) questionnaire in an adolescent sample. *American Journal on Addictions*. 2013; 22:60–66. [PubMed: 23398228]
- Pagano ME, Krentzman AR, Onder CC, Baryak JL, Murphy JL, Zywiak WH, Stout RL. Service to Others in Sobriety (SOS). *Alcoholism Treatment Quarterly*. 2010; 28:111–127. [10.1080/07347321003656425](https://doi.org/10.1080/07347321003656425) [PubMed: 21399711]
- Pagano ME, White WL, Kelly JF, Stout RL, Tonigan JS. The 10-year course of alcoholics anonymous participation and long-term outcomes: A follow-up study of outpatient subjects in Project MATCH. *Substance Abuse*. 2013; 34:51–59. [PubMed: 23327504]
- Passetti LL, Godley SH. Adolescent substance abuse treatment clinicians' self-help meeting referral practices and adolescent attendance rates. *Journal of Psychoactive Drugs*. 2008; 40:29–40.
- Passetti LL, Godley SH, Godley MD. Youth participation in mutual support groups: History, current knowledge, and areas for future research. *Journal of Groups in Addiction & Recovery*. 2012; 7:253–278.
- Rasch, G. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche; 1960.
- Reckase MD. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*. 1979; 4:207–230. <http://dx.doi.org/10.2307/1164671>.
- Scheuneman JD, Subhiyah RG. Evidence for the validity of a Rasch model technique for identifying differential item functioning. *Journal of Outcome Measurement*. 1998; 2:33–42. [PubMed: 9661730]
- Scott CK, Dennis ML, Foss MA. Utilizing recovery management checkups to shorten the cycle of relapse, treatment reentry, and recovery. *Drug & Alcohol Dependence*. 2005; 78:325–338. [PubMed: 15893164]
- Scott CK, Foss MA, Dennis ML. Pathways in the relapse—treatment—recovery cycle over 3 years. *Journal of Substance Abuse Treatment*. 2005; 28:S63–S72. <http://dx.doi.org/10.1016/j.jsat.2004.09.006>. [PubMed: 15797640]
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company; 2002.
- Smith EV. Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*. 2002; 3:205–231. [PubMed: 12011501]

- Stewart DG, Brown SA. Withdrawal and dependency symptoms among adolescent alcohol and drug abusers. *Addiction Research & Theory*. 1995; 90:627–635.
- Substance Abuse and Mental Health Services Administration. Results from the 2013 National Survey on Drug Use and Health: Summary of national findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2014. Retrieved from www.samhsa.gov/data/sites/default/files/NSDUHResultsPDFWHTML2013/Web/NSDUHresults2013.pdf
- Tims FM, Dennis ML, Hamilton N, Buchan BJ, Diamond GS, Funk R, Brantley LB. Characteristics and problems of 600 adolescent cannabis abusers in outpatient treatment. *Addiction*. 2002; 97:S46–S57.
- Titus JC, Smith DC, Dennis ML, Ives M, Twanow L, White MK. Impact of a training and certification program on the quality of interviewer-collected self-report assessment data. *Journal of Substance Abuse Treatment*. 2012; 42:201–212. <http://dx.doi.org/10.1016/j.jsat.2011.10.017>.
- Tonigan JS, Connors GJ, Miller WR. Alcoholics Anonymous Involvement Scale (AAI) Scale: Reliability and norms. *Psychology of Addictive Behaviors*. 1996; 10:75–80.
- Tonigan JS, Toscova R, Miller WR. Meta-analysis of the literature on alcoholics anonymous: Sample and study characteristics moderate findings. *Journal of Studies on Alcohol and Drugs*. 1996; 57:65–67.
- U.S. Department of Health and Human Services. ASPE research brief. 2013. Retrieved from http://aspe.hhs.gov/health/reports/2013/mental/rb_mental.cfm
- U.S. Department of Justice. The economic impact of illicit drug use on American society. Washington DC: Author; 2011. Retrieved from www.justice.gov/archive/ndic/pubs44/44731/44731p.pdf
- U.S. Executive Office of the President. National drug control strategy. 2013. Retrieved from <http://www.whitehouse.gov/ondcp/national-drug-control-strategy>
- Wang WC. Modeling effects of differential item functioning in polytomous items. *Journal of Applied Measurement*. 2000; 1:63–82. [PubMed: 12023558]
- Weiss RD, Griffin ML, Najavits LM, Hufford C, Kogan J, Thompson HJ, Siqueland L. Self-help activities in cocaine dependent patients entering treatment: Results from the NIDA collaborative cocaine treatment study. *Drug and Alcohol Dependence*. 1996; 43:79–86. [http://dx.doi.org/10.1016/S0376-8716\(96\)01292-6](http://dx.doi.org/10.1016/S0376-8716(96)01292-6). [PubMed: 8957146]
- Wilson, M. *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2005.
- Wright, BD.; Stone, MH. *Best test design*. Chicago, IL: University of Chicago, MESA Press; 1979.

Biographies

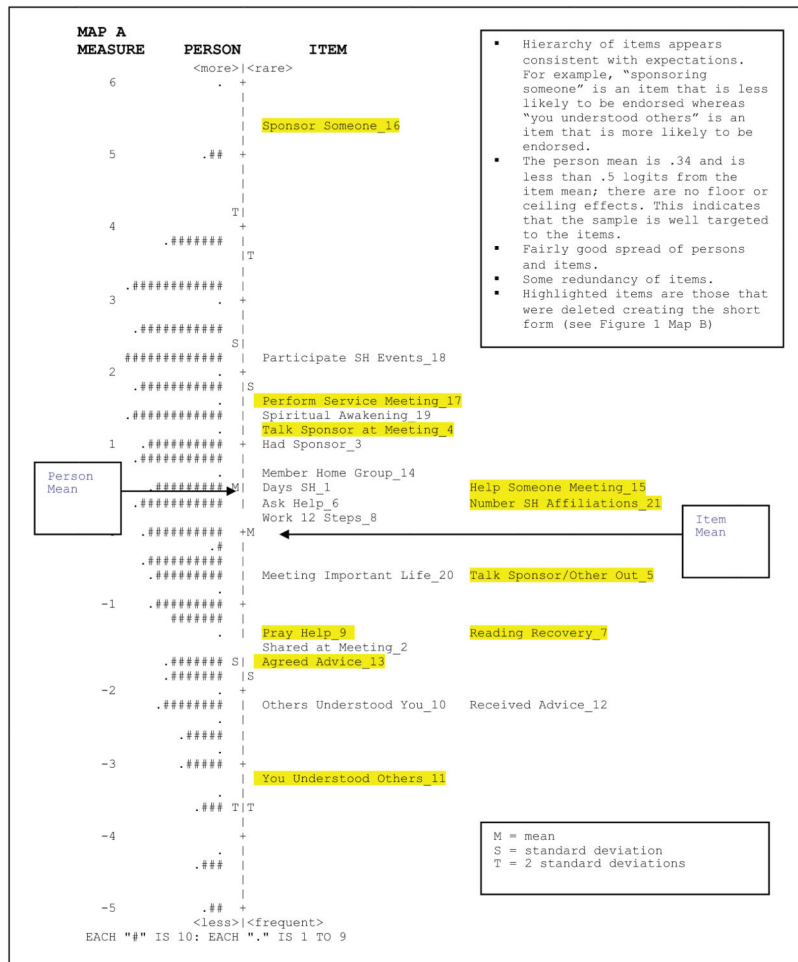
Karen M. Conrad, PhD, RN, MPH, is a research associate professor in Health Policy and Administration in the School of Public Health at the University of Illinois at Chicago and co-owner of Program Metrics LLC in Oak Park, Illinois. Her academic work has focused on occupational health. She has published a number of papers using the Rasch measurement model.

Kendon J. Conrad, PhD, is Professor Emeritus in Health Policy and Administration in the School of Public Health at the University of Illinois at Chicago and co-owner of Program Metrics LLC in Oak Park, Illinois. His work has focused on the measurement of constructs useful in research and evaluation of novel programs in substance abuse, mental health, and aging. Recently he has worked to develop a decision support system to investigate and ameliorate elder abuse.

Lora L. Passetti, MS, is a Research Projects Manager at Chestnut Health Systems. She has worked on several substance abuse treatment studies of adolescents and adults in outpatient and residential settings. Her research interest is adolescent participation in 12-step programs.

Rodney R. Funk is a research associate at the Lighthouse Institute of Chestnut Health Systems in Normal, Illinois. He holds SPSS's highest certification for data mining methods and modeling. His research interests are in the analysis of complex hierarchical and intensive data analytic models, as well as using clustering techniques to interpret heterogeneous samples and trends over time.

Michael L. Dennis, PhD, is a senior research psychologist at the Lighthouse Institute and director of the GAIN Coordinating Center at Chestnut Health Systems in Normal, Illinois. His research also focuses on understanding and predicting how people move through cycles of substance use, and crime, treatment, incarceration and periods of recovery, as well as how to experimentally test strategies for improving recovery management over time. He has published widely on recovery management, integrating clinical and research assessment, measurement, intensive data collection with smartphones, and evaluation research.



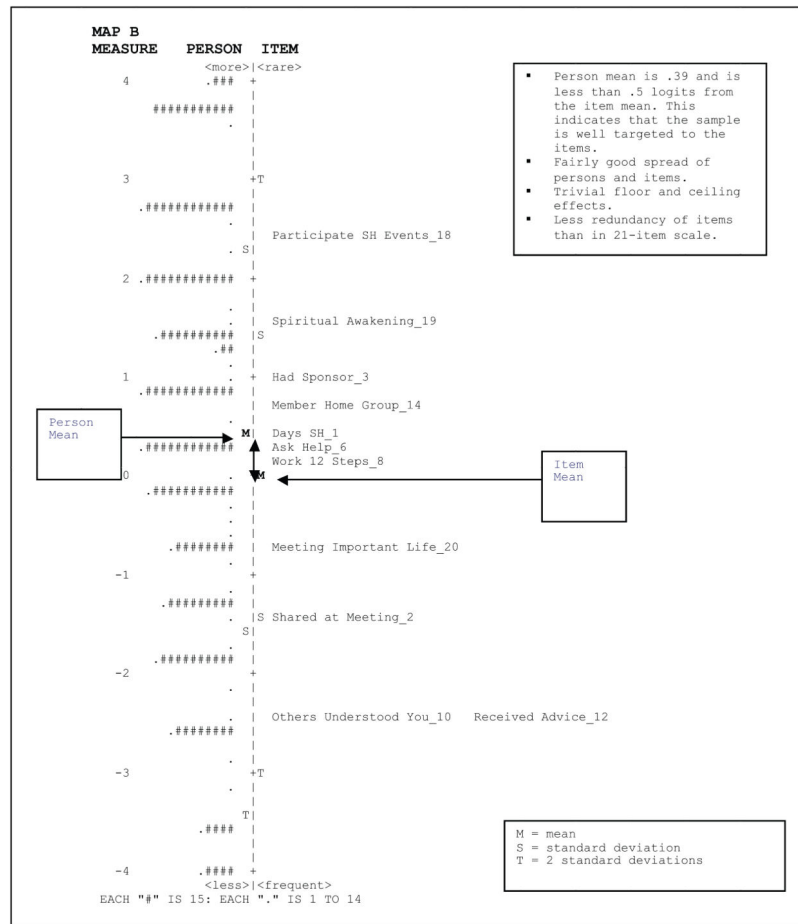


Figure 1. Self-help Involvement Scale Wright maps simultaneously displaying item difficulty (endorsability) hierarchy and person ability (level of self-help involvement). Map A: full 21 item version. Map B: 11-item short form version. The 11 items highlighted in Map A are those deleted in the short-form. Explanatory details of Wright Map are provided in the text boxes.

Table 1
 Items, Item Measures, and Item Fit Statistics for the Full 21-item and 11-item Short Form Self-help Involvement Scale.

Item Number	Question	Item Label/Number	SHIS 21			SHIS 11 Short Form		
			Item Measure	Infit	Outfit	Item Measure	Infit	Outfit
1	"During the past 90 days, on how many days have you attended one or more self-help group meetings (such as AA, NA, CA, or Social Recovery) for your alcohol or other drug use?"	Days SH_1	0.38	1.38	1.40	0.47	1.31	1.30
2	In the past 90 days, have you Spoken up (shared) during a self-help meeting?	Shared at Meeting_2	-1.35	1.00	0.92	-1.39	0.98	1.03
3	Had a sponsor?	Had Sponsor_3	0.93	0.77	0.60	1.02	0.84	0.70
4	Talked to your sponsor at a meeting?	Talk Sponsor at Meeting_4	1.19	0.77	0.57			
5	Talked to your sponsor or other members outside of a meeting?	Talk Sponsor/Other Out_5	-0.63	0.88	0.72			
6	Asked for help from your sponsor or another member?	Ask Help_6	0.26	0.81	0.72	0.32	0.85	0.75
7	Read the Twelve Steps and Twelve Traditions or other recovery related readings?	Reading Recovery_7	-1.38	0.95	0.70			
8	Actively worked the 12 steps?	Work 12 Steps_8	0.11	0.99	0.95	0.16	1.01	0.97
9	Prayed or meditated for help from your Higher Power?	Pray Help_9	-1.4	1.14	1.35			
10	Felt that other people in the meeting understood you and your problems?	Others Understood You_10	-2.29	0.84	0.64	-2.4	0.91	0.81
11	Felt that you understood other people in the meeting and their problems?	You Understood Others_11	-3.16	0.85	0.75			
12	Gotten advice or ideas about how to handle your problems better from a meeting or meeting members?	Received Advice_12	-2.26	0.80	0.57	-2.37	0.89	0.70
13	Agreed with the advice you were given?	Agreed Advice_13	-1.63	0.93	0.98			
14	Considered yourself a member of a home group?	Member Home Group_14	0.68	1.05	1.02	0.76	1.04	1.03
15	Helped someone else from a meeting?	Help Someone Meeting_15	0.3	0.98	0.92			
16	Sponsored someone else?	Sponsor Someone_16	5.33	1.11	1.77			
17	Performed a service like setting up for meetings, serving as meeting chairperson, treasurer, or literature person?	Perform Service Meeting_17	1.65	0.96	0.87			
18	Participated in conferences, dances, picnics, or other social activities sponsored by a self-help group?	Participate SH Events_18	2.29	0.99	1.25	2.48	1.04	1.45
19	Had a spiritual awakening through meeting, working the steps, or reading 12-step related literature?	Spiritual Awakening_19	1.47	0.99	0.87	1.6	1.01	0.92
20	Considered participation in self-help meetings an important part of your life?	Meeting Important Life_20	-0.66	0.89	0.73	-0.66	0.93	0.79
21	"Do you consider yourself to be a 'regular attendee or member' of any specific 12 step fellowships, faith-based, secular or other recovery groups? (Which groups?) (Any other groups?) See list	Number SH Affiliations_21	0.17	1.51	1.58			

Note. SHIS = Self-Help Involvement Scale; AA = Alcoholics Anonymous; NA = Narcotics Anonymous; CA = Cocaine Anonymous.