

## Item banks for substance use from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Severity of use and positive appeal of use<sup>☆</sup>



Paul A. Pilkonis<sup>a,\*</sup>, Lan Yu<sup>a,b</sup>, Nathan E. Dodds<sup>a</sup>, Kelly L. Johnston<sup>a</sup>,  
Suzanne M. Lawrence<sup>a</sup>, Thomas F. Hilton<sup>c,1</sup>, Dennis C. Daley<sup>a</sup>, Ashwin A. Patkar<sup>d,e</sup>,  
Dennis McCarty<sup>f</sup>

<sup>a</sup> Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, United States

<sup>b</sup> Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, United States

<sup>c</sup> NIDA, Indian Harbour Beach, FL 32937, United States

<sup>d</sup> Department of Psychiatry, Duke University School of Medicine, Durham, NC 27705, United States

<sup>e</sup> Department of Community and Family Medicine, Duke University School of Medicine, Durham, NC 27705, United States

<sup>f</sup> Public Health and Preventive Medicine, Oregon Health and Science University, Portland, OR 97239, United States

### ARTICLE INFO

#### Article history:

Received 30 June 2015

Received in revised form

10 September 2015

Accepted 10 September 2015

Available online 21 September 2015

#### Keywords:

Substance use

Drug use

Item response theory

Measurement

### ABSTRACT

**Background:** Two item banks for substance use were developed as part of the Patient-Reported Outcomes Measurement Information System (PROMIS®): severity of substance use and positive appeal of substance use.

**Methods:** Qualitative item analysis (including focus groups, cognitive interviewing, expert review, and item revision) reduced an initial pool of more than 5300 items for substance use to 119 items included in field testing. Items were written in a first-person, past-tense format, with 5 response options reflecting frequency or severity. Both 30-day and 3-month time frames were tested. The calibration sample of 1336 respondents included 875 individuals from the general population (ascertained through an internet panel) and 461 patients from addiction treatment centers participating in the National Drug Abuse Treatment Clinical Trials Network.

**Results:** Final banks of 37 and 18 items were calibrated for severity of substance use and positive appeal of substance use, respectively, using the two-parameter graded response model from item response theory (IRT). Initial calibrations were similar for the 30-day and 3-month time frames, and final calibrations used data combined across the time frames, making the items applicable with either interval. Seven-item static short forms were also developed from each item bank.

**Conclusions:** Test information curves showed that the PROMIS item banks provided substantial information in a broad range of severity, making them suitable for treatment, observational, and epidemiological research in both clinical and community settings.

© 2015 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

The Patient-Reported Outcomes Measurement Information System (PROMIS®) is designed to improve assessment of self-reported

<sup>☆</sup> Supplementary materials for this article can be found by accessing the online version of this paper.

\* Corresponding author at: Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213, United States.

E-mail address: [pilkonispa@upmc.edu](mailto:pilkonispa@upmc.edu) (P.A. Pilkonis).

<sup>1</sup> Retired.

outcomes using state-of-the-art psychometric methods (see [www.nihpromis.org](http://www.nihpromis.org)). PROMIS item banks provide a comprehensive profile of health status, including physical functioning, pain, fatigue, sleep disturbance, emotional distress, alcohol use, and social participation (Buysse et al., 2010; Cella et al., 2010; Fries et al., 2009; Pilkonis et al., 2011, 2013; Revicki et al., 2009). PROMIS is the most ambitious attempt to date to apply models from item response theory (IRT) to health-related assessment (Cella et al., 2010; Hilton, 2011; Reeve et al., 2007). We report here on the development and calibration of two item banks measuring severity of substance use and the positive appeal of substance use, an important motivational factor influencing the development and

treatment of substance use disorder (Cox et al., 2015; Dow and Kelly, 2013).

In the area of substance use, analyses relying on IRT have been used for multiple purposes: investigation of the dimensionality and informational value of diagnostic criteria for substance use disorders (Langenbucher et al., 2004; Lynskey and Agrawal, 2007; Saha et al., 2012); refinement of existing instruments (e.g., the Alcohol, Smoking and Substance Involvement Screening Test, ASSIST; the Global Appraisal of Individual Needs, GAIN), often to shorten them for purposes of more efficient screening (Ali et al., 2013; Stucky et al., 2014); and development of scales to assess constructs relevant to substance use, e.g., risk behavior related to injecting drugs, alexithymia (Hendryx et al., 1992; Janulis, 2014). The DSM-5 Substance-Related Disorders Work Group reviewed 39 IRT studies on the scaling and calibration of criteria for substance abuse and dependence. On the basis of this evidence, the Work Group eliminated the distinction between abuse and dependence in favor of a single category—substance use disorder—operationalized with 11 criteria (Hasin et al., 2013). The diagnostic criteria are sufficiently unidimensional for calibration with IRT models, but they are high-threshold items most relevant at the severe end of the continuum of substance use (Hartman et al., 2008; Langenbucher et al., 2004; Lynskey and Agrawal, 2007). Our goal was to identify not only items consistent with the diagnostic criteria but also items that were more normally distributed in a sample that included drug users in the community as well as drug users seeking treatment. Such items will provide more information across a broader range of the continuum of substance use. For this reason, they should constitute more sensitive measures of treatment outcome and result in a single metric of severity that can be used across treatment, observational, and epidemiological studies.

Drugs of abuse are too various to encourage the development of separate item banks for individual substances, especially in the PROMIS context where the mandate is to develop item banks for generic constructs relevant to all chronic diseases. Therefore, we took a broad approach and created items that referred to “drugs” in general. This approach is also consistent with the clinical epidemiology of many secondary and tertiary care facilities where patients appearing for treatment often abuse multiple substances (Stinson et al., 2005).

Substance use may vary over time as a function of availability of different drugs, resources for procuring drugs, periods of incarceration, and other life circumstances. Thus, an important issue is the preferred time frame for items. A shorter time frame (e.g., the past 30 days) is preferable while patients are in treatment and during the early stages of recovery when they are at high risk of relapse. Thus, a shorter period would be most suitable for clinical trials, treatment outcome studies, and comparative effectiveness studies where the priorities are repeated measurements and responsiveness to change over relatively brief periods. By contrast, a longer time frame (e.g., the past 3 months) allows aggregation of use over a more extended interval and may be more appropriate for observational or epidemiological studies. Rather than making an a priori decision about time frame, we chose to investigate empirically the psychometric implications of 30-day and 3-month time frames. Thus, the same item pool was administered to two different samples receiving one or the other of the two time frames, allowing examination of the performance of items under the two conditions.

## 2. Methods

### 2.1. Development of item pool

**2.1.1. Comprehensive literature searches.** We performed searches in the MEDLINE, PsycINFO, and Health and Psychosocial Instruments (HaPI) databases. Details of the methodology are reported in Klem et al. (2009). The searches generated 3388 abstracts linked to more than 80 unique measures of substance use. Copies of the

measures were gathered from both electronic and print sources, and measures were reviewed at the item level.

**2.1.2. Conceptual organization of items.** The initial substance use item pool contained over 5300 items. We organized the items into conceptually meaningful categories using a hierarchical approach informed by previous empirical work (e.g., factor analyses) and clinical formulation. Prior work had divided substance use items into domains relevant both to the DSM categorization of substance use disorders, e.g., syndromal indicators of dependence, negative consequences associated with substance use (Green et al., 2011; Krueger et al., 2004; Muthén, 2006; Saha et al., 2007), and to broader themes, e.g., motivations for drug use, drug-seeking behaviors (Jones et al., 2001; Korcha et al., 2011; Pabst et al., 2009). Our hierarchical structure for substance use included five domains: motivations for use, intended effects of use, drug-seeking behaviors, patterns of consumption, and negative consequences of use. We also created 38 distinct facets across the 5 domains. For example, within the domain of negative consequences, we included separate facets for physical, mental, social, and legal consequences.

**2.1.3. Focus groups.** To ensure comprehensive coverage of the substance use domain from the patient perspective, we conducted two focus groups and performed thematic analyses of the topics discussed (Castel et al., 2008; Kelly et al., 2011). Members of the groups ( $n = 15$ ) were recruited from outpatient substance use treatment programs. The average age was 41 ( $SD = 9$ ). Females (53%) and members of minority groups were well-represented (race = African American 53%, ethnicity = Hispanic 6%). Two-thirds of the participants (67%) had no formal education beyond high school.

Using semi-structured scripts, facilitators prompted participants to discuss their experiences with substance use and the characteristics of problematic consumption. Research staff reviewed process notes from the groups and audio recordings, paying close attention to appraisals of use (consequences of substance use, intended versus actual effects of use) and contexts and triggers of substance-related experiences. Participants discussed a variety of consequences of substance use: physical (e.g., pain, contracting sexually transmitted diseases), mental (e.g., decreased concentration, emotional distress), social (e.g., loss of friends and relationships), and financial (e.g., job loss, debt). In an era when legalization of marijuana for medical and recreational purposes is becoming more common, we were also attentive to perceived positive aspects of substance use. The appeal of substance use included both increasing positive emotions (e.g., feeling happy and social) and alleviating negative emotions (e.g., reducing depression and anxiety). The general goal was to enrich our item pool with content not always represented on traditional questionnaires. We were also eager to identify potential items with lower thresholds for endorsement (e.g., used only on weekends, used only in certain situations).

**2.1.4. Cognitive interviews.** Forty-three participants from outpatient substance use treatment programs were recruited for cognitive interviews. All items were reviewed by at least six individuals. An effort was made to include adequate numbers of women and members of minority groups (47% and 53%, respectively). Forty percent reported no formal education beyond high school. An interviewer met with participants and asked each to “think aloud” while responding to items. Participants were then prompted for feedback on the language and clarity of items and the relevance of the content. Of the 338 items reviewed by participants, 72% were acceptable as presented, 15% were re-written, and 14% were deleted. Of the re-written items, a majority (82%) required a revision to the item stem (content) versus the time frame or response options.

The cognitive interviews and continued revision of the items by members of the research team produced a pool of 119 items for general substance use. More details of the qualitative review process and its results are described in the supplement available online.<sup>2</sup>

### 2.2. Sampling

We administered the items to an Internet sample of 875 participants from the general population and a clinical sample of 461 patients in treatment for substance use disorders in North Carolina, Oregon, and Pennsylvania. The treatment centers participated in the National Drug Abuse Treatment Clinical Trials Network (McCarty et al., 2008; Tai et al., 2010). The internet sample was collected through YouGov, a web-based polling firm in Palo Alto, CA. Eligible participants reported use of an illicit substance within either the past 30 days or the past 3 months in initial screening. This procedure resulted in two calibration samples: substance use in the past 30 days ( $n = 765$ ; internet = 459, clinical = 306), and substance use in the past 3 months ( $n = 571$ ; internet = 416, clinical = 155).

Demographic characteristics of the YouGov sample and the clinical sample (combined across all sites) are summarized in Table 1. Members of the clinical sample were younger ( $p < .001$ ) and less well educated ( $p < .001$ ) than participants from the community. There were racial differences between the two samples ( $p < .001$ ),

<sup>2</sup> Supplementary material can be found by accessing the online version of this paper.

**Table 1**  
Demographic characteristics of the calibration sample.

Characteristic	Clinical sample ( <i>n</i> = 461) %	Community sample ( <i>n</i> = 875) %
Sex		
Male	50.8	49.6
Ethnicity		
Hispanic	5.4	13.1
Race		
American Indian/Alaska Native	2.2	0.0
Asian	0.9	1.3
Black/African American	22.3	10.3
White	68.8	82.4
Other/Multiracial	5.9	6.1
Education		
High school diploma or less	53.4	41.0
Further educational attainment	46.6	59.0
Mean Age ( <i>SD</i> )	37.3 (11.4)	44.6 (16.5)

due largely to more African American participants and fewer white participants in the clinical sample. However, there were fewer Hispanic participants in the clinical sample ( $p < .001$ ), and there was no difference in sex ratio.

### 2.3. Measures

The substance use item pool used in field testing contained 119 items assessing motivations for use (17 items), intended effects of use (30 items), drug-seeking behaviors (9 items), patterns of consumption (25 items), and negative consequences of use (38 items). Participants also completed the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST v3.0; WHO ASSIST Working Group, 2002), a commonly used “legacy” instrument. In addition, they reported the ages at which they first used alcohol and drugs, and they provided a medical history that asked about lifetime diagnoses and functional impairment in nine areas (e.g., cardiovascular problems, cancer, mental health).

### 2.4. Data analysis

**2.4.1. General strategy.** Our major goal was to identify the most robust latent constructs within the item pool and to document sufficient unidimensionality for each of them to proceed with IRT analyses. First, we inspected frequency distributions of individual items for sparse cells. We then investigated dimensionality by dividing the samples randomly into two subsamples, one for exploratory factor analysis (unweighted least squares EFA,  $n = 276$  for the 3-month time frame and  $n = 413$  for the 30-day time frame) and the second for confirmatory factor analysis (CFA,  $n = 295$  for the 3-month time frame and  $n = 352$  for the 30-day time frame). The EFAs and CFAs were conducted using Mplus 6.1 with promax rotation (Muthén and Muthén, 2010). In the CFAs, the items were treated as categorical variables, and the robust weighted least squares estimator was used.

**2.4.2. Item response theory analysis.** The 2-parameter graded response model (GRM; Samejima, 1969) is the most commonly used IRT model for polytomous items. The GRM has a slope parameter and  $n - 1$  threshold parameters for each item, where  $n$  is the number of response categories (5 in the present analyses). The slope parameter measures item discrimination, i.e., how well the item differentiates between higher versus lower levels of severity of substance use. Informative items have large slope parameters. Threshold parameters measure item difficulty, i.e., the ease versus difficulty of endorsing different response options for an item. Items were calibrated using IRTPRO v2.1 (Cai et al., 2011).

**Table 2**  
Substance use reported on the ASSIST.

Substance	Ever used		Used in past 3 months	
	Clinical sample ( <i>n</i> = 461) %	Community sample ( <i>n</i> = 875) %	Clinical sample ( <i>n</i> = 461) %	Community sample ( <i>n</i> = 875) %
Cannabis	93.5	76.5	64.2	57.9
Cocaine	83.1	38.1	39.8	7.1
Opioids	75.2	23.7	65.7	10.0
Amphetamines	58.8	36.2	30.6	8.3
Hallucinogens	50.7	38.1	4.6	4.5
Sedatives	50.1	35.0	31.7	19.0
Inhalants	14.6	11.0	1.5	2.6
Tobacco	92.0	75.1	86.3	45.4
Alcohol	94.1	95.6	72.2	85.4

Differential item functioning (DIF) occurs when characteristics (e.g., age, gender, or ethnicity) which may seem extraneous to the assessment of the constructs under consideration affect measurement. An item is identified as functioning differentially if the item is more (or less) difficult to endorse or more (or less) discriminating in one group (e.g., men) compared to a reference group (e.g., women) when the different subgroups have been matched on the latent trait under investigation. With regard to demographic characteristics, we conducted DIF analyses (for both uniform and non-uniform DIF) on the basis of age, gender, and education (high school education or less versus further educational attainment). We focused on these variables because the relevant comparison groups were adequately represented. We also used a median split on age (less than 43 versus 43 or older) to compare younger and older respondents. Other potential comparison groups (e.g., white versus non-white respondents, Hispanic versus non-Hispanic respondents) were less equally divided. In addition, we conducted DIF analyses on the basis of severity of overall substance use (using a median split on the total substance involvement scores computed from the ASSIST). This analysis examined the performance of items based on differential exposure to drugs and documented their consistency across the full spectrum of substance use. Two different DIF procedures were employed—the IRT likelihood ratio method (Thissen et al., 1993) embedded in IRTPRO and an ordinal logistic regression procedure (Zumbo, 1999). Items were considered for removal if they showed significant DIF ( $p < .01$ ) by both methods (Teresi et al., 2009).

## 3. Results

### 3.1. Drug, alcohol, and tobacco use in the two samples

Table 2 summarizes the lifetime use of substances reported on the ASSIST. The most frequently used drugs in the clinical sample were cannabis (94%), cocaine (83%), and opioids (75%), and in the community sample, cannabis (77%), cocaine (38%), and hallucinogens (38%). For the past 3 months, the clinical sample reported using opioids (66%), cannabis (64%), and cocaine (40%) most frequently; in the community sample, the comparable percentages were cannabis (58%), sedatives (19%), and opioids (10%). The samples also differed in median age of first substance use (clinical sample = 14, community sample = 17) and the percentage using prior to age 15 (clinical sample = 50%, community sample = 26%), a commonly cited risk factor (Grant and Dawson, 1998; King and Chassin, 2007; Swendsen et al., 2009). The lifetime prevalence of injecting any drug was 39% in the clinical sample and 11% in the community sample (see Table 3).

Alcohol use in the past three months was common in both samples: 72% of the clinical sample and 85% of the community sample. Finally, 86% of the clinical sample were current smokers versus 45% for the community sample. High-risk behaviors and medical consequences associated with such behaviors are summarized in Table 3. In summary, the levels and patterns of substance use differed in the two samples (as expected), enhancing the generalizability of the item banks across clinical and community populations.

### 3.2. Frequency distributions

Within the initial pool of 119 items, there were no items with response categories having less than 1% response. There were 7 items (6%) having at least one response category with between 1%

**Table 3**  
High-risk behaviors and medical status.

	Clinical sample ( <i>n</i> = 461)	Community sample ( <i>n</i> = 875)
<i>In the past 3 months, while using drugs...<sup>a</sup></i>	%	%
Did risky things	62.0	7.6
Had unprotected sex	28.4	6.0
Used sex to get drugs	21.5	4.4
Taken advantage of sexually	9.8	2.4
<i>Lifetime</i>		
Injected a substance	39.0	10.6
Used someone else's needle	15.0	5.6
HIV/AIDS	1.1	2.7
Sexually transmitted disease	25.8	15.2
Liver problems	23.4	7.1

<sup>a</sup> The response options for these items were “never,” “rarely,” “sometimes,” “often,” or “almost always.” The percentages reflect endorsements at the level of “sometimes” or higher.

and 3% response. However, the sparse cells for all 7 items had at least 18 respondents, ranging from 18 to 39. Therefore, we retained all 5 response categories for all items for further analyses.

### 3.3. Factor analyses

**3.3.1. Exploratory factor analyses.** The initial EFA of the 119 items with a 3-month time frame (*n* = 276) yielded 10 factors with eigenvalues greater than 1, and the initial EFA with a 30-day time frame (*n* = 413) yielded 11 factors with eigenvalues greater than 1. In both analyses, however, the first factor (with eigenvalues of 76.7 for 3-months and 78.0 for 30-days) dominated, with ratios of the first to second eigenvalue of 9.92 (3 months) and 13.1 (30 days). In all analyses, this first factor included a large subset of items with high factor loadings reflecting many aspects of substance use, and it appeared to be a general indicator of the severity of use.

Inspection of 1- through 5-factor solutions for the 3-month data supported a 2-factor solution as most interpretable, with items on the second factor reflecting recreational use of drugs, pleasurable aspects of the experience, and perceived positive consequences of drug use. Additional factors were less interpretable and appeared to represent “splinters” (i.e., small subsets of items with modest loadings on a primary factor and often with cross-loadings on other factors). Inspection of 1- through 5-factor solutions for the 30-day data supported a 3-factor solution as most interpretable, with the second and third factors including the items regarding positive aspects of drug use that loaded significantly on factor 2 with the 3-month data. Taken together, the EFAs documented the importance of 2 subsets of items: 95 items best understood as measuring the severity of substance use, and 22 items reflecting the positive appeal of substance use.

A second round of EFAs examined the two subsets of items separately in both the 3-month and 30-day data. Despite the large number of items in the severity subset, a 1-factor solution was best supported with both time frames. The eigenvalue for the first factor was 68.5 in the 3-month data and 69.3 in the 30-day data, producing ratios of the first to second eigenvalue of 21.2 in the 3-month data and 26.7 in the 30-day data. We retained the entire pool of 95 items for CFA and IRT calibration, all of which had factor loadings greater than .60 in the repeat EFAs. The repeat EFAs with the subset of 22 items reflecting positive appeal of substance use also supported a 1-factor solution. In this case, however, we chose to delete 2 items with smaller factor loadings (less than .50).

**3.3.2. Confirmatory factor analyses.** We performed single-factor CFAs on the reduced item pools, using the second half of the samples (3-month time frame, *n* = 295; 30-day time frame, *n* = 352). For

the pool of 95 general severity items, there was excellent evidence of unidimensionality. For the 3-month data, all factor loadings were greater than .62, and several fit indices were uniformly strong: CFI = .982, TLI = .982, and RMSEA = .042. For the 30-day data, the results were similar: all factor loadings greater than .55, CFI = .981, TLI = .980, and RMSEA = .044. The smaller pool of 20 items reflecting positive appeal of substance use was less homogeneous in CFA terms (reflected primarily in a larger RMSEA), but the results were adequate to document that responses to these items were largely a function of a single underlying latent construct. For the 3-month data, all factor loadings were greater than .47, and the fit indices were CFI = .938, TLI = .928, and RMSEA = .146. For the 30-day data, the results were again similar: all factor loadings greater than .50, CFI = .939, TLI = .929, and RMSEA = .149.

### 3.4. IRT calibrations

The two item pools—95 items for severity of substance use and 20 items for positive appeal of substance use—were calibrated separately, and the calibrations were done twice, first with the 3-month data and then with the 30-day data, using the two-parameter GRM. IRT modeling of the 95-item severity pool with both time frames revealed many locally dependent item pairs, suggesting considerable redundancy among the items. To manage this issue, we identified 19 “core” items. These items were selected using both clinical and patient perspectives. From a clinical perspective, we included items reflecting the diagnostic criteria for substance use disorder (i.e., commonly accepted markers of the presence and severity of substance use), and from the patient perspective, we included items reflecting themes that were prominent in our focus groups (e.g., consequences of substance use, intended versus actual effects of use). After these core items were selected, we added back 34 other items that were locally independent from the 19 core items, resulting in a new set of 53 items. We repeated the IRT calibrations with these items, again using the 3-month and 30-day data separately. The item parameter estimates, test information curves, and standard error estimates were very similar for the two time frames. We computed intraclass correlation coefficients (ICCs) between the time frames for the discrimination parameters and the four location parameters. These ICCs were large: 0.95 for the discrimination parameter, and 0.98, 0.98, 0.97, and 0.97, respectively, for the four location parameters. Given these results, we combined the 3-month and 30-day data to compute final, definitive IRT item parameters. For the combined sample, the *n* was 1336. From the pool of 53 severity items, we identified 16 items that were locally dependent following this final calibration, leaving an item bank of 37 items.

The IRT results with the pool of 20 items reflecting the positive appeal of substance use also showed that separate calibrations with the 3-month and 30-day data were similar. The ICCs for the item parameters between the two time frames were 0.96 for the discrimination parameter and 0.99 for all four location parameters. Therefore, we again used the combined data to compute final IRT item parameters. Calibration with the combined data revealed 2 items that were locally dependent, and they were removed from the bank, leaving 18 items for the final version.

DIF analyses for both item banks showed that the IRTPRO and logistic regression methods did not jointly identify any items displaying DIF on the basis of age, gender, education, or more versus less severe substance use; thus, no items were eliminated for this reason. The Flesch–Kincaid readability test was performed on the 55 items in the two banks. The grade level for the 37 severity items in the aggregate was grade 2.5; for the 18 positive appeal items, it was grade 2.3. Tables 4 and 5 summarize the item banks,

**Table 4**  
Calibrated items: severity of substance use.

Item Stem	Slope (Discrimination)	Location Thresholds			
		Level 1	Level 2	Level 3	Level 4
I felt that my drug use was out of control*	6.56	0.47	0.71	0.93	1.11
My desire to use drugs seemed overpowering*	6.19	0.35	0.57	0.94	1.24
Drugs were the only thing I could think about*	5.83	0.45	0.74	1.08	1.39
I felt I needed help for my drug problem	5.65	0.42	0.61	0.77	0.97
My drug use caused problems with people close to me*	5.28	0.39	0.60	0.90	1.21
I spent more time using drugs than I intended	4.97	0.37	0.64	0.91	1.21
I have a drug problem*	4.97	0.19	0.46	0.68	0.90
I did whatever I had to do to use drugs	4.68	0.57	0.87	1.17	1.43
Using drugs caused problems for me	4.65	0.36	0.61	0.91	1.21
I am addicted to drugs	4.65	0.18	0.47	0.66	0.93
I sold my belongings to buy drugs	4.58	0.80	0.99	1.28	1.56
I craved drugs*	4.52	0.11	0.40	0.77	1.18
Someone told me that I should cut down on my drug use	4.49	0.40	0.67	1.00	1.33
My drug use kept me from getting things done that I needed to do	4.35	0.24	0.54	0.95	1.29
I used drugs when I felt guilty	4.19	0.51	0.74	1.16	1.45
I needed to use more drugs to get the same effect I used to get	4.18	0.30	0.55	0.89	1.23
I spent a lot of time using drugs*	4.05	0.11	0.49	0.89	1.20
When I used drugs, I failed to take care of myself	3.88	0.52	0.82	1.21	1.56
When I used drugs, I experienced emotional problems	3.79	0.36	0.66	1.08	1.51
I planned my activities around using drugs	3.64	0.31	0.59	1.05	1.41
I used drugs to avoid my responsibilities	3.53	0.65	0.90	1.33	1.65
I did risky things while using drugs	3.49	0.45	0.78	1.28	1.67
I did things that could have gotten me into trouble while using drugs	3.37	0.40	0.68	1.18	1.60
When I used drugs, I felt alone	3.17	0.44	0.76	1.23	1.64
I used drugs to feel "normal"	2.98	0.25	0.46	0.90	1.24
I bought drugs for other people, but kept them for my own use	2.98	0.97	1.23	1.58	1.94
When I stopped using drugs, I got headaches	2.95	0.54	0.84	1.35	1.75
I almost died because of my drug use	2.91	1.26	1.55	1.89	2.34
I saved my money to buy drugs	2.83	0.35	0.60	1.11	1.54
When I used drugs, I got into fights	2.78	0.88	1.27	1.73	2.07
When I used drugs, I hid my drug use from others	2.77	0.13	0.41	0.82	1.16
When I used drugs, I felt anxious	2.66	0.29	0.74	1.29	1.72
I used drugs to reward myself	2.64	0.18	0.51	1.08	1.54
I used drugs because nearly all the people I know use drugs	2.35	0.40	0.80	1.28	1.76
I used drugs to feel more alert	2.29	0.46	0.77	1.37	1.85
I used drugs because people pressured me to use	2.08	1.14	1.69	2.34	2.80
I drove a car while under the influence of drugs	1.80	0.53	0.93	1.45	2.14

Note. Items are rank-ordered on the basis of their slope (discrimination) parameters. Items included in the short form are marked with an asterisk.

**Table 5**  
Calibrated items: positive appeal of drug use.

Item stem	Slope (discrimination)	Location thresholds			
		Level 1	Level 2	Level 3	Level 4
I used drugs to feel more confident*	6.30	0.45	0.65	0.96	1.24
I used drugs to feel good about myself*	5.58	0.43	0.63	0.95	1.23
I used drugs to make it easier to talk to people*	5.12	0.45	0.63	1.02	1.35
Drugs made me feel like I could do anything*	4.94	0.50	0.71	1.02	1.27
I used drugs to feel close to someone	4.01	0.63	0.85	1.26	1.61
I used drugs to change my mood*	3.81	-0.05	0.16	0.69	1.08
While using drugs, I liked myself better	3.66	0.46	0.74	1.11	1.42
I used drugs to have a good time*	3.47	-0.18	0.10	0.61	1.04
I used drugs to be a part of something	3.41	0.67	0.96	1.40	1.72
I used drugs to be open to new experiences	3.38	0.45	0.77	1.25	1.58
I used drugs to be more creative	3.32	0.40	0.66	1.19	1.61
People seemed to like me better when I used drugs	3.21	0.57	0.90	1.33	1.74
I used drugs because I liked the feeling*	2.92	-0.49	-0.16	0.23	0.64
I used drugs to relax	2.87	-0.37	-0.11	0.46	1.01
I used drugs while enjoying TV, music, or video games	2.71	-0.14	0.11	0.62	1.07
I enjoyed using drugs	2.51	-0.54	-0.13	0.35	0.83
I used drugs recreationally	2.22	-0.38	-0.01	0.59	1.04
I used drugs when I was with other people	2.17	-0.36	0.06	0.72	1.24

Note. Items are rank-ordered on the basis of their slope (discrimination) parameters. Items included in the short form are marked with an asterisk.

together with their IRT parameters. Figs. 1 and 2 display the test information curves (and plots of corresponding standard errors). Standard errors of .30 correspond approximately to classical test theory reliabilities of .90. At this threshold, the effective range of measurement for the severity of use item bank was about  $-.6$  to  $+2.7$  SDs and for the positive appeal of substance use item bank, about  $-.8$  to  $+2.2$  SDs.

### 3.5. Selection of items for short forms

For some applications where CAT is not feasible, static short forms may be a useful alternative. To develop short forms, we rank ordered all severity of use and positive appeal of use items on four criteria: discrimination parameters, the percentage of times the item would have been selected in a simulated CAT for each

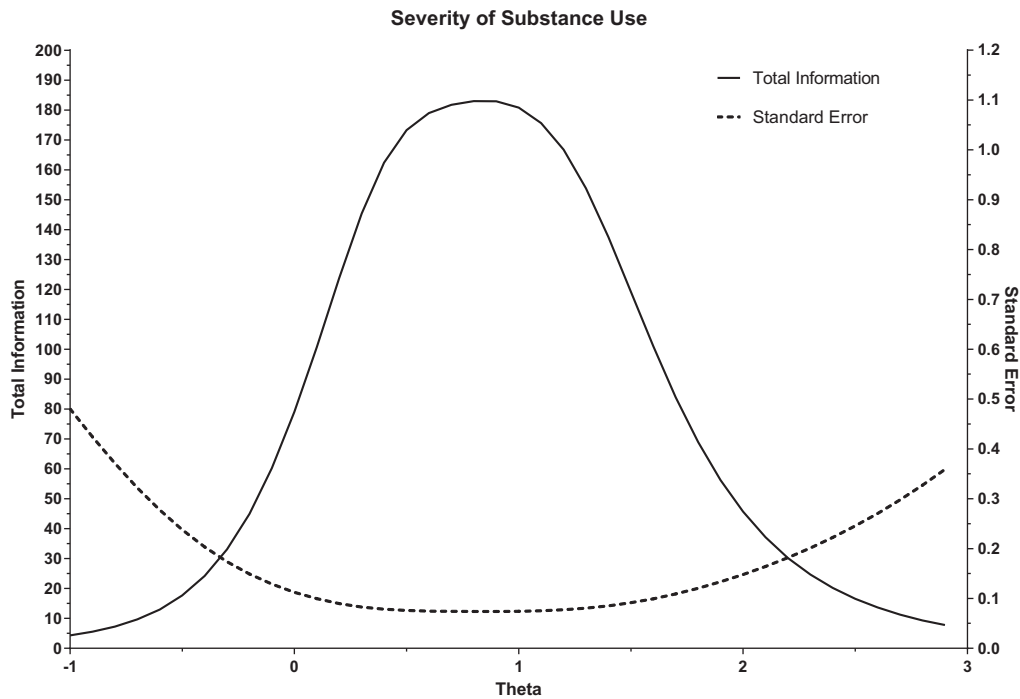


Fig. 1. Test information and standard error curves for the severity of substance use item bank.

item bank based on the observed data from our calibration sample, expected information under the standard normal distribution with a mean of 0 and SD of 1, and expected information under a normal distribution with a larger SD, i.e., a mean of 0 and SD of 1.5 (Choi et al., 2010). The CAT simulations were performed using the Firestar program (Choi, 2009). For the CAT simulations, we set the minimum number of items to be administered to 8 and the maximum number of items to be administered to be the full bank. We

selected 7 items for each short form based on the convergence of the four psychometric criteria, the content of candidate items, and location parameters (i.e., we tried to include some items with lower thresholds to increase the precision of the short forms closer to the floor). In Tables 4 and 5, asterisks identify the items selected for the short forms.

The internal consistency of the short forms was excellent. Alpha coefficients were .94 for the severity of use short form and .92

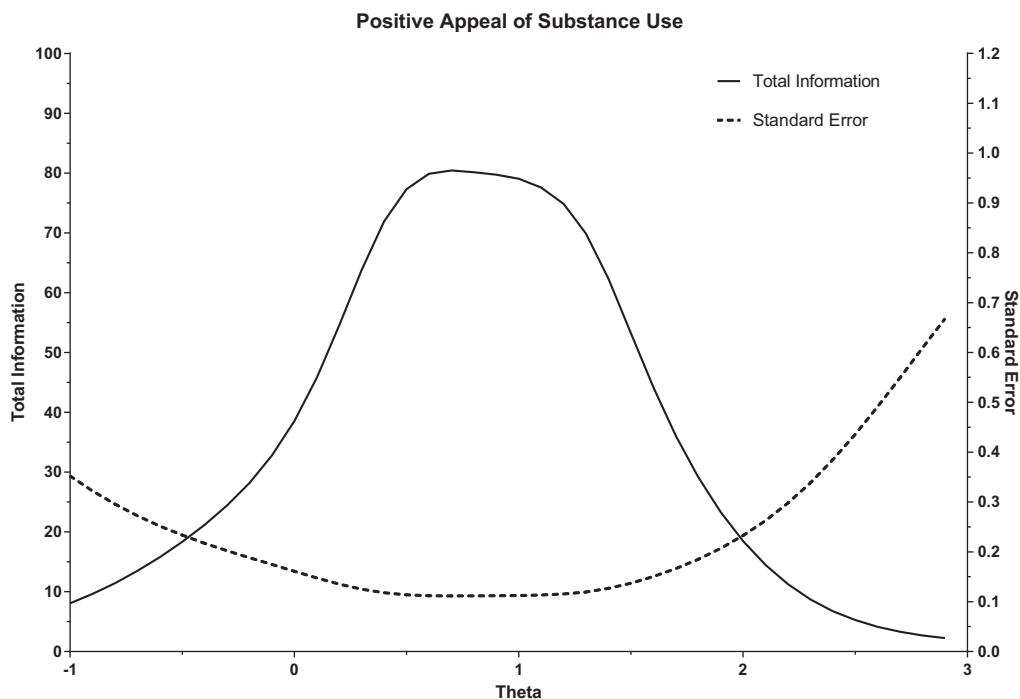


Fig. 2. Test information and standard error curves for the positive appeal of substance use item bank.

**Table 6**  
DSM-5 criteria and corresponding severity of substance use items.

DSM-5 Criterion
<i>Severity of substance use item</i>
A1. Uses more than intended, or for longer than intended <i>I spent more time using drugs than I intended</i>
A2. Efforts to control or cut back on use have been unsuccessful <i>I felt that my drug use was out of control*</i>
A3. Large amounts of time are spent obtaining, using, or recovering from use <i>I spent a lot of time using drugs*</i>
A4. Cravings (the presence of a strong desire to use) <i>My desire to use drugs seemed overpowering*</i> <i>Drugs were the only thing I could think about*</i> <i>I craved drugs*</i>
A5. Recurrent use resulting in problems at work, home, or school <i>Using drugs caused problems for me</i> <i>My drug use kept me from getting things done that I needed to do</i>
A6. Continued use despite recurrent social or interpersonal problems resulting from use <i>My drug use caused problems with people close to me*</i> <i>When I used drugs, I got into fights</i> <i>When I used drugs, I hid my drug use from others</i>
A7. Curtailing important activities in favor of use <i>I planned my activities around using drugs</i> <i>I used drugs to avoid my responsibilities</i>
A8. Use despite potentially hazardous outcomes <i>I did risky things while using drugs</i> <i>I did things that could have gotten me into trouble while using drugs</i> <i>I almost died because of my drug use</i> <i>I drove a car while under the influence of drugs</i>
A9. Continued use despite its causing or exacerbating a persistent physical or psychological problem <i>When I used drugs, I failed to take care of myself</i> <i>When I used drugs, I experienced emotional problems</i> <i>When I used drugs, I felt alone</i> <i>When I used drugs, I felt anxious</i>
A10. Tolerance or a need for increased amounts of substance <i>I needed to use more drugs to get the same effect I used to get</i>
A11. Withdrawal symptoms <i>When I stopped using drugs, I got headaches</i>

Note. Items included in the short form are marked with an asterisk.

for the positive appeal of substance use short form. Correlations between the theta scores derived from the short forms and the corresponding full item banks were large: .94 for severity of use and .96 for positive appeal of use.

### 3.6. Convergence with DSM-5 diagnostic criteria

Inspection of the 37 items in the severity item bank revealed that 23 of them could be mapped onto the 11 diagnostic criteria for substance use disorder now included in the DSM-5 (see Table 6). Each criterion was represented by at least one item (up to a maximum of four). The 14 items distinct from the diagnostic criteria reflected a variety of other themes: acknowledgment of having a drug problem, ways of procuring drugs, emotional triggers and consequences of substance use, and social motives and pressures to use drugs.

## 4. Discussion

The mixed (qualitative and quantitative) PROMIS methodology was used to create two item banks relevant to substance use: a 37-item bank measuring severity of substance use and an 18-item bank measuring positive appeal of substance use. As expected, we found an overall severity factor (and item bank). However, the item bank for positive appeal of use is more novel. This second item bank for positive appeal may be especially relevant for community samples, and it is timely because legalization of marijuana (for both medical and recreational purposes) is increasingly common. In the calibration data from the community sample, cannabis was the most commonly used substance (current and lifetime), and it

generated the largest ASSIST SI correlation with the theta score for positive appeal of use,  $r = .61$  (as well as the severity theta score,  $r = .56$ ). In general, there were stronger relationships in the community sample between SI scores for all substances and positive appeal of use.

Given its content, the item bank for severity of use could be used for the development of a short form that captures the DSM-5 diagnostic criteria for substance use disorder (see Table 6). The PROMIS mandate is to create measures of severity (not diagnosis) that can be used as common metrics across all health states and diseases. However, it is typical for investigators and clinicians to ask about the relevance of PROMIS measures for screening and diagnostic purposes and about thresholds on the severity metric that should motivate clinical concern or intervention. A short form focused on diagnostic criteria might be useful clinically. At the same time, it must be emphasized that the mapping of the PROMIS items to the DSM-5 criteria is preliminary and that the validity of the items in relation to both the individual DSM-5 criteria and the presence of a diagnosis requires further investigation. Five of the criteria are linked to only one PROMIS item, and “physical problems” in criterion A9 are not well represented.

In general, there is a need for further studies of the validity of the new item banks. Data from the calibration sample provided only preliminary evidence of validity, e.g., correlations with ASSIST SI scores (see supplement available online<sup>3</sup>). Validity studies can take many forms, but consistent with past PROMIS precedents (Pilkonis et al., 2014) they should include (a) psychometric studies comparing the operating characteristics of the PROMIS item banks and short forms with other commonly used legacy measures and (b) longitudinal studies of responsiveness to change in circumstances where one would (or would not) expect variability over time.

In this context, it is important to note that analysis of longitudinal data regarding substance use presents special challenges. Before asking about severity of substance use, the item banks require a screening question to document the presence of some use: “In the past 30 days, have you used drugs (other than alcohol or your prescribed medications)?” Thus, the item banks provide information about two related but different phenomena: exposure to drugs and severity of use when exposed. Analyzing such data requires novel models, and an approach we have found useful is the two-part semi-continuous model (Olsen and Schafer, 2001). The model allows simultaneous analysis of binary (yes/no) and continuous (severity) data, yielding parameter estimates (intercepts and slopes) for functions that model both exposure and severity. Other discussions of analyzing data in which information about severity is contingent on the presence of some behavior or event are also available (e.g., Liu and Verkuilen, 2013; Reardon and Raudenbush, 2006). It is important to acknowledge that there is only a single screening question for the new item banks and that the use of a single question may limit reliability and validity. At the same time, there are many legacy instruments for assessing exposure to drugs, including items regarding the types of drugs used and the frequency and quantity of use. We used the ASSIST for just this purpose, and other investigators may also wish to include companion measures that provide more detailed information regarding exposure to drugs. It should also be noted that abuse of prescription pain medication (obtained either medically or illicitly) is an increasing public health problem. A separate PROMIS item bank has been created for this purpose, and a description of its development and calibration will be the subject of a separate report.

In summary, the development of two new item banks for substance use—severity of use and positive appeal of use—adds to

<sup>3</sup> Supplementary material can be found by accessing the online version of this paper.

the existing body of PROMIS measures. The initial psychometric characteristics of the item banks support their use as CATs or short forms, with either version providing brief, precise, and efficient measures relevant to both clinical and community samples. Further studies of the validity of the item banks are now appropriate to support their adoption.

## Author disclosures

### Role of funding source

PROMIS<sup>®</sup> was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD (deceased), and David Scott Tulskey, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztejn, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James Witter, MD, PhD. An award from the National Institute on Drug Abuse (UG1 DA015815) provided additional support for Dennis McCarty's participation. This article uses data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Government or PROMIS. See [www.nihpromis.org](http://www.nihpromis.org) for additional information on the PROMIS<sup>®</sup> initiative.

## Contributors

Paul A. Pilkonis, PhD, contributed to study conception, design, and implementation and took responsibility for drafting the manuscript. Lan Yu, PhD, provided data analysis and interpretation. Nathan Dodds, BS, Kelly L. Johnston, MPH, Suzanne Lawrence, MA, Thomas F. Hilton, PhD, Dennis C. Daley, PhD, Ashwin A. Patkar, MD and Dennis McCarty, PhD contributed to study implementation (literature reviews, conceptual organization of items, item and intellectual property reviews) and data collection (focus groups, cognitive interviews, coordination of field testing for calibration). All authors reviewed and approved the final manuscript.

## Conflict of interest

There are no conflicts of interest for any authors.

## Acknowledgements

We would also like to thank our collaborators at the recruitment sites: Center for Psychiatric and Chemical Dependency Services, Pittsburgh, PA (Dorothy J. Sandstrom, MS; Janis McDonald; Trey Ghee, MSW); CODA, Inc., Portland, OR (Katharina Wiest, PhD; Rosalie Gordon, BA; Kasie Cloud, MSW), Coastal Horizons Center, Inc., Wilmington, NC (Kenny S. G. House, BA, LCAS, CCS; Michael Younkens, BS, BBA), and Southlight Healthcare, Raleigh, NC (Alyssa H. Kalata, PhD, Scott Luetgenau, BSW, Reynolds (Tad) Clodfelter, Jr., PsyD).

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.drugalcdep.2015.09.008>.

## References

- Ali, R., Meena, S., Eastwood, B., Richards, I., Marsden, J., 2013. Ultra-rapid screening for substance-use disorders: the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST-Lite). *Drug Alcohol Depend.* 132, 352–361.
- Buysse, D., Yu, L., Moul, D.E., Germain, A., Stover, A., Dodds, N.E., Johnston, K.L., Shablesky-Cade, M.A., Pilkonis, P.A., 2010. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep* 33, 781–792.
- Cai, L., Thissen, D., du Toit, S.H.C., 2011. IRTPRO: Flexible, Multidimensional, Multiple Categorical IRT Modeling [Computer Software]. Scientific Software International, Lincolnwood IL.
- Castel, L.D., Williams, K.A., Bosworth, H.B., Eisen, S.V., Hahn, E.A., Irwin, D.E., Kelly, M.A.R., Morse, J., Stover, A., DeWalt, D.A., DeVellis, R.F., 2008. Content validity in the PROMIS social health domain: a qualitative analysis of focus group data. *Qual. Life Res.* 17, 737–749.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B.B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., DeVellis, R., DeWalt, D., Fries, J.F., Gershon, R., Hahn, E.A., Lai, J.S., Pilkonis, P., Revicki, D., Rose, M., Weinfurt, K., Hays, R., 2010. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item bank: 2005–2008. *J. Clin. Epidemiol.* 63, 1179–1194.
- Choi, S., 2009. Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Appl. Psychol. Meas.* 33, 644–645.
- Choi, S.W., Reise, S.P., Pilkonis, P.A., Hays, R.D., Cella, D., 2010. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual. Life Res.* 19, 125–136.
- Cox, W.M., Klinger, E., Fardadi, J.S., 2015. The motivational basis of cognitive determinants of addictive behaviors. *Addict. Behav.* 44, 16–22.
- Dow, S.J., Kelly, J.F., 2013. Listening to youth: adolescents' reasons for substance use as a unique predictor of treatment response and outcome. *Psychol. Addict. Behav.* 27, 1122–1131.
- Fries, J.F., Cella, D., Rose, M., Krishnan, E., Bruce, B., 2009. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J. Rheumatol.* 36, 2061–2066.
- Grant, B.F., Dawson, D.A., 1998. Age of onset of drug use and its association with DSM-IV drug abuse and dependence: results from the national longitudinal alcohol epidemiologic survey. *J. Subst. Abuse* 10, 163–173.
- Green, B.A., Ahmed, A.O., Marcus, D.K., Walters, G.D., 2011. The latent structure of alcohol use pathology in an epidemiological sample. *J. Psychiatr. Res.* 45, 225–233.
- Hartman, C.A., Gerlhorn, H., Crowley, T., Sakai, J.T., Stallings, M., Young, S.E., Rhee, S.H., Corley, R., Hewitt, J.K., Hopfer, C.J., 2008. Item response theory analysis of DSM-IV cannabis abuse and dependence criteria in adolescents. *J. Am. Acad. Child. Adolesc. Psychiatry* 47, 165–173.
- Hasin, D.S., O'Brien, C.P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Compton, W.M., Crowley, T., Ling, W., Petry, N.M., Schuckit, M., Grant, B.F., 2013. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am. J. Psychiatry* 170, 834–851.
- Hendryx, M.S., Haviland, M.G., Gibbons, R.D., Clark, D.C., 1992. An application of item response theory to alexithymia assessment among abstinent alcoholics. *J. Pers. Assess.* 58, 506–515.
- Hilton, T.F., 2011. The promise of PROMIS<sup>®</sup> for addiction. *Drug Alcohol Depend.* 119, 229–234.
- Janulis, P., 2014. Improving measurement of injection drug risk behavior using item response theory. *Am. J. Drug Alcohol Abuse* 40, 143–150.



- Jones, B.T., Corbin, W., Fromme, K., 2001. A review of expectancy theory and alcohol consumption. *Addiction* 96, 57–72.
- Kelly, M.A., Morse, J.Q., Stover, A., Hofkens, T., Huisman, E., Shulman, S., Eisen, S.V., Becker, S.J., Weinfurt, K., Boland, E., Pilkonis, P.A., 2011. Describing depression: congruence between patient experiences and clinical assessments. *Br. J. Clin. Psychol.* 50, 46–66.
- King, K.M., Chassin, L., 2007. A prospective study of the effects of age of initiation of alcohol and drug use on young adult substance dependence. *J. Stud. Alcohol Drugs* 68, 256–265.
- Klem, M.L., Saghafi, E., Abromitis, R., Stover, A., Dew, M.A., Pilkonis, P.A., 2009. Building PROMIS item banks: librarians as co-investigators. *Qual. Life Res.* 18, 881–888.
- Korcha, R.A., Polcin, D.L., Bond, J.C., Lapp, W.M., Galloway, G., Pharm, D., 2011. Substance use and motivation: a longitudinal perspective. *Am. J. Drug Alcohol Abuse* 37, 48–53.
- Krueger, R.F., Nichol, P.E., Hicks, B.M., Markon, K.E., Patrick, C.J., Iacono, W.G., McGue, M., 2004. Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychol. Assess.* 16, 107–119.
- Langenbucher, J.W., Labouvie, E., Martin, C.S., Sanjuan, P.M., Bavly, L., Kirisci, L., Chung, T., 2004. An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. *J. Abnorm. Psychol.* 113, 72–80.
- Liu, Y., Verkuilen, J., 2013. Item response modeling of presence-severity items: application to measurement of patient-reported outcomes. *Appl. Psychol. Meas.* 37, 58–75.
- Lynskey, M.T., Agrawal, A., 2007. Psychometric properties of DSM assessments of illicit drug abuse and dependence: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Psychol. Med.* 37, 1345–1355.
- McCarty, D., Fuller, B., Kaskutas, L.A., Wendt, W.W., Nunes, E.V., Miller, M., Forman, R., Magruder, K.M., Arfken, C., Copersino, M., Floyd, A., Sindelar, J., Edmondson, E., 2008. Treatment programs in the National Drug Abuse Treatment Clinical Trials Network. *Drug Alcohol Depend.* 92, 200–207.
- Muthén, B., 2006. Should substance use disorders be considered as categorical or dimensional? *Addiction* 101, 6–16.
- Muthén, L.K., Muthén, B., 2010. *Mplus User's Guide (Version 6.2)*. Muthén & Muthén, Los Angeles, CA.
- Olsen, M.K., Schafer, J.L., 2001. A two-part random effects model for semicontinuous longitudinal data. *J. Am. Stat. Assoc.* 96, 730–745.
- Pabst, A., Baumeister, S.E., Kraus, L., 2009. Alcohol-expectancy dimensions and alcohol consumption at different ages in the general population. *J. Stud. Alcohol Drugs* 71, 46–53.
- Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., Riley, W.T., Cella, D., 2011. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, anger. *Assessment* 18, 263–283.
- Pilkonis, P.A., Yu, L., Colditz, J., Dodds, N.E., Johnston, K.L., Maihoefer, C., Stover, A.M., Daley, D.C., McCarty, D., 2013. Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS®): use, consequences, and expectancies. *Drug Alcohol Depend.* 130, 167–177.
- Pilkonis, P.A., Yu, L., Dodds, N.E., Johnston, K.L., Maihoefer, C., Lawrence, S.M., 2014. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS®) in a three-month observational study. *J. Psychiatr. Res.* 56, 112–119.
- Reardon, S., Raudenbush, S.W., 2006. A partial independence item response model for surveys in which responses to filter questions determine whether subsequent questions are asked. *Sociol. Methodol.* 36, 257–300.
- Reeve, B.B., Hays, R.D., Bjorner, J.B., Cook, K.F., Crane, P.K., Teresi, J.A., Thissen, D., Revicki, D.A., Weiss, D.J., Hambleton, R.K., Liu, H., Gershon, R., Reise, S.P., Cella, D., 2007. Psychometric evaluation and calibration of Health-Related Quality of Life item banks: plans for the Patient-Reported Outcome Measurement Information System (PROMIS). *Med. Care* 45, S22–S31.
- Revicki, D., Chen, W., Harnam, N., Cook, K., Amtmann, D., Callahan, L.F., Jensen, M.P., Keefe, F.J., 2009. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain* 146, 158–169.
- Saha, T., Compton, W.M., Chou, S.P., Smith, S., Ruan, W.J., Huang, B., Pickering, R.P., Grant, B.F., 2012. Analyses related to the development of DSM-5 criteria for substance use related disorders: 1. Toward amphetamine, cocaine and prescription drug use disorder continua using item response theory. *Drug Alcohol Depend.* 122, 38–46.
- Saha, T.D., Stinson, F.S., Grant, B.F., 2007. The role of alcohol consumption in future classification of alcohol use disorders. *Drug Alcohol Depend.* 89, 82–92.
- Samejima, F., 1969. Estimation of Latent Ability Using a Response Pattern of Graded Scores (Psychometrika Monograph No. 17). Psychometric Society, Richmond, VA.
- Stinson, F.S., Grant, B.F., Dawson, D.A., Ruan, W.J., Huang, B., Saha, T., 2005. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug Alcohol Depend.* 80, 105–116.
- Stucky, B.D., Edelen, M.O., Ramchand, R., 2014. A psychometric assessment of the GAIN Individual Severity Scale (GAIN-GISS) and Short Screeners (GAIN-SS) among adolescents in outpatient treatment programs. *J. Subst. Abuse Treat.* 46, 165–173.
- Swendsen, J., Conway, K.P., Degenhardt, L., Dierker, L., Glantz, M., Jin, R., Merikangas, K.R., Sampson, N., Kessler, R.C., 2009. Socio-demographic risk factors for alcohol and drug dependence: the 10-year follow-up of the national comorbidity survey. *Addiction* 104, 1346–1355.
- Tai, B., Straus, M.M., Liu, D., Sparenborg, S., Jackson, R., McCarty, D., 2010. The first decade of the National Drug Abuse Treatment Clinical Trials Network: bridging the gap between research practice to improve drug abuse treatment. *J. Subst. Abuse Treat.* 38, S4–S13.
- Teresi, J.A., Ocepek-Welickson, K., Kleinman, M., Eimicke, J.P., Crane, P.K., Jones, R.N., Lai, J.S., Choi, S.W., Hays, R.D., Reeve, B.B., Reise, S.P., Pilkonis, P.A., Cella, D., 2009. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): an item response theory approach. *Psychol. Sci. Q.* 51, 148–180.
- Thissen, D., Steinberg, L., Wainer, H., 1993. Detection of differential item functioning using the parameters of item response models. In: Holland, P.W., Wainer, H. (Eds.), *Differential Item Functioning*. Lawrence Erlbaum, Hillsdale, NJ, pp. 67–113.
- WHO ASSIST Working Group, 2002. The Alcohol Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction* 97, 1183–1194.
- Zumbo, B.D., 1999. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Directorate of Human Resources Research and Education, Department of National Defense, Ottawa, ON.