



Development and psychometric validation of the Clinical Assessment Interview for Negative Symptoms (CAINS)

William P. Horan ^{a,*}, Ann M. Kring ^b, Raquel E. Gur ^c, Steven P. Reise ^d, Jack J. Blanchard ^e

^a VA Greater Los Angeles Healthcare System, University of California, Los Angeles, California, United States

^b University of California, Berkeley, CA, United States

^c University of Pennsylvania, United States

^d University of California, Los Angeles, California, United States

^e University of Maryland, College Park, United States

ARTICLE INFO

Article history:

Received 4 May 2011

Received in revised form 29 June 2011

Accepted 30 June 2011

Available online 27 July 2011

Keywords:

Schizophrenia

Rating scale

Expression

Experience

Negative symptoms

Clinical trials

ABSTRACT

Progress in the development of new pharmacological and psychosocial treatments for the negative symptoms of schizophrenia is impeded by limitations of available assessment instruments. The multi-site Collaboration to Advance Negative Symptom Assessment in Schizophrenia (CANSAS) was established to develop and validate a new clinical rating scale using a transparent, iterative, and data-driven process. The Clinical Assessment Interview for Negative Symptoms (CAINS) was designed to address limitations of existing measures and assess consensus-based sub-domains, including asociality, avolition, anhedonia, affective blunting, and alolia. The structure and psychometric properties of the CAINS were evaluated in a sample of 281 schizophrenia and schizoaffective outpatients at four sites. Converging structural analyses indicated that the scale was comprised of two moderately correlated factors – one reflecting experiential impairments (diminished motivation and enjoyment of social, vocational, and recreational activities) and one reflecting expressive impairments (diminished non-verbal and verbal communication). Item-level analyses revealed generally good distributional properties, inter-rater agreement, discriminating anchor points, and preliminary convergent and discriminant validity. Results indicate that the CAINS is a promising new measure for quantifying negative symptoms in clinical neuroscience and treatment studies. Results guided item modification or deletion, and the reliability and validity of the revised, shorter version of the CAINS is in the final phase of development within the CANSAS project.

Published by Elsevier B.V.

1. Introduction

Negative symptoms substantially impede functional recovery for people with schizophrenia (Kirkpatrick et al., 2006). Despite their clinical significance, current treatments do not adequately address negative symptoms – there are not yet any medications with a specific indication for negative symptoms and psychosocial interventions show similarly limited benefits (Leucht et al., 1999; Montgomery and Zwieten-Boot, 2006). To address this critical treatment need, the NIMH-Negative Symptom Consensus Development Conference (Kirkpatrick et al., 2006) recommended that the field required a new negative symptoms measure that can be productively used in pharmacological trials. The Collaboration to Advance Negative Symptom Assessment of Schizophrenia (CANSAS) was established to develop and validate a “next-generation” clinical rating scale by following data-driven, iterative, and transparent process (Blanchard et al., 2011). This report describes the psychometric

evaluation of a beta version of the Clinical Assessment Interview for Negative Symptoms (CAINS) in a large outpatient sample. This effort is unlike any other scale development project to date in that it includes a large and diverse patient population and adopts a comprehensive empirical approach to item generation, selection, and retention.

The CAINS was designed to address limitations of existing instruments (Horan et al., 2006; Blanchard et al., 2011) and assess the five consensus negative symptom sub-domains (Kirkpatrick et al., 2006). Ratings of asociality, avolition, and anhedonia are based on interviewees' reported subjective experiences of motivation and emotion, as well as frequency of actual engagement in relevant activities. Asociality assesses the degree to which close social bonds are valued and desired, and frequency of social interactions. Avolition assesses level of interest and motivation, and initiation and persistence of behavior. Anhedonia assesses experience and frequency of consummatory pleasure and anticipatory pleasure. The final two domains are rated based on observable behaviors throughout the interview. Blunted affect ratings also include prompts to elicit positive and negative emotions. Ratings for alolia include measures of speech output.

A feasibility study of an early version of the scale (Forbes et al., 2010) demonstrated good internal consistency and inter-rater agreement, and

* Corresponding author at: UCLA Department of Psychiatry & Biobehavioral Sciences, 300 UCLA Medical Plaza, Suite 2255, Los Angeles, CA 90095-6968, United States. Tel.: +1 310 206 8181; fax: +1 310 206 3651.

E-mail address: horan@ucla.edu (W.P. Horan).

very good convergent and discriminant validity with other symptom and functional outcome measures. However, several areas needed refinement, including skewed and restricted range of anhedonia items, low inter-item correlations in asociality and avolition domains, marginal inter-rater agreement for items in alogia and avolition domains, and difficulties distinguishing among anchor points for several items.

We report here a comprehensive psychometric evaluation of the revised CAINS in a large, diverse sample of outpatients with schizophrenia or schizoaffective disorder based on recommendations that clinically stable patients are preferred for negative symptom treatment development studies (Kirkpatrick et al., 2006; Laughren and Levin, 2006). The first goal was to examine the scale's latent structure through a series of complementary structural analyses. Clarification of the underlying structure of the scale items, which were designed to comprehensively cover five consensus-based content domains, is critical for optimal assessment of the negative symptoms (Blanchard and Cohen, 2006; Blanchard et al., 2011). This was followed by a series of scale development analyses, including analyses of item- and scale-level characteristics, within- and between-site inter-rater agreement, and a preliminary analysis of discriminant and convergent validity. The over-arching goal was to refine the CAINS for use in the final scale development phase of the CANSAS project.

2. Methods and materials

2.1. Participants

Participants were 281 people with schizophrenia ($n=223$) or schizoaffective disorder ($n=58$), ages 18–60, recruited from outpatient clinics at the four CANSAS sites (UCLA, UC-Berkeley, University of Pennsylvania, and the University of Maryland). Patients met diagnostic criteria based on the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID; First et al., 1996). Exclusion criteria were: episode of major depression or mania within the last month; substance dependence in the last six months; substance abuse in the last month; $IQ < 70$; history of head injury or neurological disorder; insufficient English fluency. All participants provided written informed consent before the study.

2.2. Procedures

A two-day training workshop preceded study initiation. Training included manual review, didactic presentations, independent ratings of videotaped CAINS and other scales interviews, and discussions. Raters were credentialed for all study instruments following practice and confirmation of competency for videotaped and in-person interviews. Procedures were identical at all sites and approved by their institutional review boards. A detailed manual for training and supervision was developed and revised throughout the course of the study.

Following revisions from the feasibility study (Forbes et al., 2010), the CAINS-beta 2 included 23 items covering (1) asociality (3 items covering family, romantic relationships, and friendships), (2) avolition (4 items covering social interactions, vocational activities, recreational activities, and self care), (3) anhedonia (9 items, covering frequency, intensity, and expected pleasure in social, physical, and recreational activities), (4) blunted affect (5 items covering facial, vocal, gestural expression as well as eye contact and spontaneous movement) and (5) alogia (2 items covering quantity of speech and spontaneous elaboration). As discussed elsewhere (Blanchard et al., 2011), the first three domains are assessed based on patients' reports of motivation, interest, and emotional experience, as well patients' reports of actual engagement in relevant social, vocational, and recreational activities. The items in these areas do not exclusively focus on level of functional attainment because poor functioning may reflect factors that are unrelated to negative symptoms (e.g., paranoia, anxiety, skill deficits, lack of opportunity). Thus, the items were designed to more closely measure constructs that are central to

negative symptoms (i.e., deficits in interest, motivation, affiliative desire).

The final two domains are based on behavioral observations during the interview. We opted to be over-inclusive with respect to the number of items, recognizing that our systematic data analytic approach to scale development would result in a smaller yet psychometrically sound instrument. All items were rated on a 0 to 4 scale with higher scores reflecting greater psychopathology. The time period covered by the interview was the past seven days. Expected pleasure assessed future pleasure with no specified time period.

Three additional measures were included to characterize the sample and for use in preliminary convergent/discriminant validity analyses: the *Brief Psychiatric Rating Scale* (BPRS; Overall and Gorham, 1962), 24-item version, assessing *Positive, Negative, Depression-Anxiety, and Agitation* (Kopelowicz et al., 2008); the *Calgary Depression Scale for Schizophrenia* (CDSS; Addington et al., 1996; Addington et al., 1990), evaluating depressive symptoms; the *Wechsler Test of Adult Reading* (WTAR; Wechsler, 2001), providing a reliable estimate of full-scale IQ.

The CAINS, BPRS, CDSS, and WTAR were administered in a fixed order. CAINS assessments were videotaped for supervision and evaluation of inter-rater agreement. A random subset of 10 CAINS videos from each site was independently rated by two different raters at each of the four sites to evaluate both within-site and between-site agreements on a common set of 40 interviews.

2.3. Data analysis

Complementary classical test theory (CTT) and item response theory (IRT; Embretson and Reise, 2000; Reise et al., 2005) analyses evaluated the latent structure of the scale, item response characteristics, inter-rater agreement, and preliminary convergent and discriminant validity. Although CAINS items were written to over-inclusively sample from five consensus-based negative symptoms domains, it was not our intention to adhere to an a priori factor structure. Instead, decisions about the ultimate structure and content of the scale were based on converging structural analyses of the overall scale and multiple item-level analyses.

We adopted several guiding principles for making data-driven decisions about retaining, modifying, or deleting items. Items that failed to meet one of these criteria were considered for deletion: (1) item fit with the factor structure of the scale, (2) inter-rater agreement within- and between-sites, (3) minimal redundancy with other items, (4) item-total correlations, (5) item skew, and (6) convergent and discriminant validity.

Data analyses thus included multiple procedures. The latent structure of the scale was evaluated using four complementary approaches. Our primary analytic approach was exploratory factor analysis using principle axis extraction with promax rotation. Complementary structural analyses were conducted to confirm the replication of this initial structure including cluster analysis (Aldenderfer and Blashfield, 1985), Bass-Ackwards analysis (Goldberg, 2002), and Mokken Scaling analyses (Sijtsma and Molenaar, 2002). Following confirmation of the CAINS general structure, additional analyses to evaluate performance of constituent items included: (a) item-level descriptive statistics, including skewness and item-total correlations; (b) inter-rater agreement with Intraclass Correlation Coefficients (ICC's; model: two-way mixed; type: absolute agreement) at both within- and between-site; (c) correlational analyses to assess whether the CAINS subscales significantly correlated with the BPRS negative symptom subscale to assess convergent validity. Discriminant validity was assessed by examining correlations between the CAINS subscales and measures of positive and affective symptoms, which were expected to be small; (d) item response theory (IRT) analyses to assess whether each of the five within-item response options was functioning adequately, and to guide item revisions and further item deletions.

3. Results

3.1. Sample characteristics

Demographic and clinical information (Table 1) on this chronically ill sample indicate diverse gender and ethnicity, allowing for confidence in applicability to diverse populations. The sample was characterized by generally high levels of functional disability. The majority of participants were unmarried, unemployed, and receiving disability benefits. Clinical symptom ratings on the BPRS and CDSS for this outpatient sample fell in the moderate to low range. Estimated IQ was in the low average range.

3.2. Scale structure

3.2.1. Exploratory factor analyses

The scree plot for the factor analysis suggested a two or three-factor solution. Overall, the analyses favored a two-factor solution, and Table 2 presents the factor loadings for the two-factor solution. Factor 1 contained all the items from the asociality, avolition, and anhedonia domains and Factor 2 contained all the items from the affective blunting and alolia domains. The two factors were moderately correlated ($r = .39$, $p < .05$). Thus, the two factor solution broadly distinguished between

Table 1
Demographic Information (N = 281).

	Mean (SD) or percent
Age	42.51 (11.75)
Sex	
Female	33.5%
Male	66.5%
Ethnicity	
Hispanic	11.7%
Not	88.3%
Race	
White	35.2%
African American	48.8%
Asian	2.8%
American Indian	1.1%
Native Hawaiian	.4%
Multiple	11.7%
Education	
Subject	12.65 (2.27)
Father	13.27 (2.74)
Mother	13.25 (2.95)
Marital	
Presently married	5.0%
Widowed	2.5%
Divorced/separated	20.3%
Never married	72.2%
Receives disability	
Yes	76.5%
No	23.5%
Has a paying job	
Yes	17.1%
No	82.9%
Living arrangements	
Unsupervised, house	70.2%
Unsupervised, boarding house	4.0%
Supervised, halfway house	4.4%
Supervised, board and care	19.6%
Transient	1.5%
Homeless	.3%
BPRS	
Thought disturbance	2.31 (1.17)
Anxiety/depression	2.20 (.91)
Anergia	1.82 (.71)
Hostility	1.74 (.81)
Agitation	1.29 (.46)
Total	42.28 (11.00)
CDSS	3.13 (6.54)
WTAR	94.68 (14.31)

Table 2
Two-factor solution for the CAINS items.

	Factor 1	Factor 2
8. Anhedonia: social frequency	.71	
4. Avolition: social	.71	
14. Anhedonia: recreation frequency	.67	
11. Anhedonia: physical frequency	.61	
9. Anhedonia: Social intensity	.58	
6. Avolition: recreation	.56	
10. Anhedonia: social expected intensity	.54	
3. Asociality: friendships	.53	
12. Anhedonia: physical intensity	.49	
1. Asociality: family	.46	
15. Anhedonia: recreation intensity	.45	
5. Avolition: vocational	.43	
16. Anhedonia: recreational expected intensity	.43	
13. Anhedonia: physical expected intensity	.41	
2. Asociality: romantic	.40	
7. Avolition: self care	.34	
23. Alogia: spontaneous		.80
22. Alogia: quantity		.78
19. Blunted: gestures		.78
18. Blunted: vocal		.78
17. Blunted: facial		.71
21. Blunted: spontaneous movement		.62
20. Blunted: eye contact		.33

Experience items (i.e., requiring participants to report on internal experience of motivation, emotion, or closeness) and Expression items (i.e., based on observations of behavior). Items that did not load above .40 on either factor included alogia: self-care and blunted affect: eye contact. The three-factor solution was more difficult to interpret, as the third factor items had extremely variable loadings and were skewed in the nonpathological direction (Table 3)¹.

We conducted additional hierarchical structural analyses including Cluster analysis, Bass-Ackwards analysis, and Mokken Scaling analysis (Supplemental On-line Materials). Convergent findings confirmed that the CAINS items do not form a “unidimensional” composite. Instead, the items break down into two homogeneous content groupings: expressive and experiential. A possible third grouping was comprised of anhedonia intensity items and eye contact and self-care items. As these items underperformed in the other analyses they were ultimately dropped.

3.2.2. Item descriptive statistics

Table 3 presents descriptive information for the CAINS items. Several items had a skew >1.0. Many of these items also had low loadings on the two factors from the structural analyses, including the anhedonia intensity items. Several items had low item-total correlations, including eye contact, spontaneous movement, and self-care.

We examined inter-item correlations to identify redundancy (On-Line Supplemental Material). The two alogia items were highly inter-correlated (.88), indicating redundancy (as in the structural analyses). The social avolition item was more strongly correlated with the asociality items than with other avolition items, suggesting that motivation for social interactions fits better with items assessing social relationships. The frequency of physical anhedonia item was somewhat redundant with the frequency of recreation anhedonia item ($r = .56$). The anhedonia intensity items were not strongly intercorrelated with one another (all but two below .35), consistent with the factor analysis.

¹ In the three-factor solution, Factor 2 was identical to the two-factor solution. Factor 1 included 11 of the 16 items from the two-factor solution Factor 1, and the remaining five items formed a third factor consisting of anhedonia intensity. Variable loadings of the third factor were: recreation past week intensity (.75) and expected intensity (.72), physical expected intensity (.59) and past week intensity (.38), and social-expected intensity (.34).

Table 3
Descriptive statistics and inter-rater agreement for CAINS items.

	Mean	SD	Skew	Item-total correlation	Within-site ICC	Between-site ICC
1. Asociality: family	1.44	1.16	.46	.45	.86	.90
2. Asociality: romantic	2.05	1.33	−.11	.44	.87	.91
3. Asociality: friendships	1.76	1.30	.28	.46	.90	.92
4. Avolition: social	1.70	1.18	.27	.68	.77	.83
5. Avolition: vocational	1.90	1.34	.04	.36	.86	.90
6. Avolition: recreation	1.47	1.10	.41	.42	.77	.83
7. Avolition: self care	.96	1.07	.76	.28	.92	.89
8. Anhedonia: social frequency	2.67	1.30	−.70	.66	.75	.80
9. Anhedonia: social intensity	1.18	1.31	1.03	.52	.94	.96
10. Anhedonia: social expected intensity	1.14	1.26	1.20	.47	.94	.97
11. Anhedonia: physical frequency	2.37	1.24	−.46	.47	.72	.71
12. Anhedonia: physical intensity	.83	1.14	1.53	.40	.84	.90
13. Anhedonia: physical expected intensity	.92	1.19	1.32	.39	.96	.98
14. Anhedonia: recreation frequency	2.48	1.21	−.59	.46	.75	.81
15. Anhedonia: recreation intensity	.83	1.07	1.43	.39	.92	.93
16. Anhedonia: recreation expected intensity	.81	1.07	1.42	.45	.95	.96
17. Blunted: facial	1.59	1.16	.14	.48	.59	.71
18. Blunted: vocal	1.14	1.07	.50	.49	.66	.69
19. Blunted: gestures	1.38	1.21	.36	.47	.57	.69
20. Blunted: eye contact	1.01	1.00	.58	.31	.60	.70
21. Blunted: spontaneous	.74	1.04	1.34	.28	.51	.63
22. Alogia: quantity	.70	1.02	1.22	.53	.62	.73
23. Alogia: spontaneous	.91	1.16	1.06	.53	.67	.73

3.2.3. Inter-rater agreement

Inter-rater agreement results are presented in the final columns of Table 3. Although agreement statistics are typically presented at the level of scale, we conducted more stringent item-level analysis to provide a rigorous assessment of item performance to inform possible item retention or revision. Both within- and between-site agreement were quite good for the Experience items. Agreement was lower, however, for the Expression items.

3.2.4. Summary of structural and item analyses

Several items failed to meet one or more criteria we established for item retention: (a) factor loading > .45 on one of the two factors; (b) inter-rater agreement > .70; (c) correlation < .55 with another item; (d) item to scale total correlation \geq .35; and (e) skew < 1.0. We thus deleted: avolition: self-care (criteria a, d); blunted: eye contact (a, b, d); blunted: spontaneous movement (d, e); all anhedonia intensity items (a, d, e); anhedonia: physical frequency (c); alogia: spontaneous elaboration (c); and avolition: social (c). The revised scale included seven Expression items and four Experience items.

The internal consistency (coefficient alpha) was .84 for the four-item Expression subscale, .77 for the seven-item Experience subscale, and .80 for all items combined. The correlation between these scales was modest ($r = .30$, $p < .001$). Within- and between-site agreements for the two subscales were above .70 (Experience subscale: within = .94 and between = .92; Expression factor: within = .77 and between = .74). The items included in these two subscales were examined in subsequent validity and IRT analyses.

3.2.5. External validity

As shown in Table 4, the Expression subscale and total scores strongly converged with BPRS negative symptom ratings; this relatively strong association is not surprising since the BPRS negative symptoms subscale focuses on observable behaviors. The correlation between the Experience subscale and BPRS negative symptoms, although significant, was small.

The CAINS subscales demonstrated good discriminant validity. The CAINS total score was not correlated with depression, positive symptoms, or IQ. The CAINS Expression scale was only modestly correlated with IQ ($r = .19$) and was not correlated with any other non-negative symptom. The CAINS Experience scale was not correlated with

IQ and had only small correlations with positive symptoms, depression (as rated on the CDSS but not BPRS), and agitation (all r s < .15).

Regarding sex and ethnicity/race, men had higher scores than women on the Expression subscale, $t(279) = 2.00$, $p = .05$, but men and women did not differ on the Experience subscale, $t(279) = 1.47$, $p > .05$. There were no effects for ethnicity or race.

3.2.6. IRT item-level analyses

IRT modeling was used to determine how clearly and meaningfully the anchor point categories (i.e., ratings of 0–4) for the retained items differentiated among individuals.² For the Expression subscale, all items were highly discriminating and, although the 5 anchor points were fairly discriminating, the spread among some anchor points was less than ideal. For example, ratings of 2 and 3 were only within a very narrow range whereas ratings of 1 and 4 tended to cover a relatively wide range of Expression symptoms (particularly for the alogia: quantity of speech item). The Experience subscale items were generally less discriminating among people than the Expression items, suggesting that these items tap into a relatively broader underlying construct. In general, the 5 anchor points were discriminating, suggesting that the response categories provide useful distinctions. However, on some items ratings of 2 and, to a lesser extent, 3 did not uniquely identify individuals on Experience symptoms. Thus, IRT analyses suggested that most anchor points for Expression and Experience items were performing reasonably well, but also provided clear guidance for revisions to better demarcate thresholds between several rating category boundaries.

4. Discussion

The goal of this collaborative project is to develop and validate a state-of-the-art clinical rating scale for negative symptoms so that the next-generation of pharmacological treatments may have more

² For each subscale, we fit a separate graded response model (GRM; Samejima, 1969) using MULTILOG (Thissen, 2003), which produces parameter estimates of a slope (i.e., how well the item discriminates among individuals) and four between-category threshold parameters (e.g., where on the latent trait continuum an individual has a 50% probability of responding 0 versus 1, 2, 3 or 4; 0, 1 versus 2, 3, 4; 0, 1, 2 versus 3, 4, and 0, 1, 2, 3 versus 4). These parameter estimates were then transformed into category response curves (CRCs), which display how the probability of responding in a particular category changes as a function of changes on the latent variable (Supplemental Analyses).

Table 4
Correlations with symptoms and estimated IQ.

	Expressive	Experiential	Total
BPRS positive	-.08	.14*	.06
BPRS negative	.72***	.23***	.52***
BPRS depression	-.03	.03	.01
BPRS agitation	-.02	.14*	.10
CDSS	.05	.12*	.11
WTAR	.19**	-.08	.04

Notes: BPRS = Brief Psychiatric Rating Scale; CDSS = Calgary Depression Scale for Schizophrenia; WTAR = Wechsler Test of Adult Reading.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

potent and clear treatment targets. Analyses of this large, diverse outpatient sample, which showed considerable functional disability, indicated that the CAINS is a very promising measure for future treatment studies in schizophrenia. Converging structural analyses indicated two underlying factors reflecting experiential and expressive negative symptoms. Item-level analyses revealed good distributional properties, inter-rater agreement, initial convergent and discriminant validity, and discrimination among individuals with anchor point categories. The results guided our final revisions of the CAINS, which we describe in more detail below.

Across a convergence of multiple analytic approaches, a two-factor model more parsimoniously defined the structure of the CAINS than the five original consensus-based negative symptom sub-domains. The first factor reflects experiential aspects of negative symptoms and includes diminished motivation for and engagement in pleasurable social, vocational, and recreational activities. The second factor reflects expressive aspects of negative symptoms and includes diminished verbal and non-verbal communicative output. Results demonstrated that the CAINS factor-derived scales showed clear convergent validity with another assessment of negative symptoms and these scales were largely independent of other non-negative symptom domains including depression and positive symptoms. This two-factor conceptualization of moderately correlated experiential and expressive negative symptoms comports with recent conceptual and empirical reviews of older interview-based measures of negative symptoms (Blanchard and Cohen, 2006; Kimhy et al., 2006; Messinger et al., 2011). Further research will determine if these factors reflect distinct neurobiological processes. Nevertheless, the convergence of our results with prior findings provides a compelling basis for organizing the CAINS into two core dimensions and crafting subscales to optimally assess each.

Based on the a priori empirical principles that guided the data analyses, the CAINS has been shortened and revised. Within the Experience domain, the most substantial changes involved anhedonia-related items. Structural and item-level analyses indicated that these items did not cohere well with other Experience items and showed substantial skew and restricted range in the non-pathological direction. Although we adopted a Likert scale approach to assess Anhedonia intensity to address similar problems from our feasibility study (Forbes et al., 2010), this was not successful. It is unclear whether these relatively non-pathological intensity scores found across studies reflect the absence of a pleasure intensity deficit, consistent with other research (Cohen and Minor, 2008; Kring and Moran, 2008) or, instead, limited sensitivity of our interview-based assessment. Nevertheless, our findings across two studies revealed psychometric limitations to rating anhedonia intensity that substantially impede their ability to demonstrate improvement in the context of a clinical trial. Thus, anhedonia intensity items were removed from the CAINS.

Anhedonia items based on the *frequency* of pleasurable experiences fared relatively better in the analyses, and thus these items were retained following revisions. First, we omitted the physical pleasure frequency item because it substantially overlapped with the recreational pleasure

frequency item and, relative to the other anhedonia frequency items, showed the lowest inter-rater agreement, greater skew toward non-pathological ratings, and was judged to be relatively less clinically important. Second, following from the IRT analyses, we adjusted the anchors for social and recreational pleasure frequency to more clearly demarcate among rating anchor points. Third, we added a provisional vocational frequency item to provide uniformity across the social, vocational, and recreational domains. Finally, we added three new provisional frequency items for expected pleasure to test the utility of assessing expected pleasure based on frequency, rather than intensity.

Results also guided changes involving avolition and asociality items. First, although self-care is included in several negative symptom scales and is widely regarded as clinically important, we removed the self-care item due to its very low coherence with other Experience items and its skew toward non-pathological ratings. Second, we removed the social avolition item due to empirical and conceptual overlap with the three asociality items. We revised the asociality items such that motivation for relationships was more clearly part of each item. Finally, IRT analyses led to modifications of probe questions and anchor points to help raters distinguish among rating points for the remaining avolition and asociality items.

Regarding the Expression items, results revealed problems with redundancy and lack of coherence for some, as well as concerns about inter-rater agreement, and demarcation among rating anchor points. Because the two alolia items were highly inter-correlated, we removed the spontaneous elaboration item, primarily because raters found the quantity of speech item simpler to use. Although eye contact and spontaneous movements are included in other scales, they showed poor coherence with other Expression items and relatively low inter-rater agreement and were thus removed. To address the lower inter-rater agreement, limited range, demarcations among rating points, and user-friendliness of these items, we revised all the remaining items to refine and simplify all rating anchor points. Finally, we added a provisional item aimed at making the format of the verbal expression (i.e., prosody) rating more consistent with the remaining Expression items.

Based on these results, we now have a substantially shorter and re-organized version of the CAINS. As shown in Table 5, the scale is now organized into two subscales that differ in content and the sources of information used to rate them. The Experience subscale consists of seven items (plus four provisional items) assessing experienced motivation and pleasure, as well as behavioral engagement in social, vocational, and recreational activities. The Expression subscale consists of four items (plus one provisional item) tapping verbal and non-verbal emotion expressive behaviors. Based on the data collected to date, the CAINS approach to rating negative symptom shows promising psychometric properties, inter-rater agreement, and convergent, discriminant validity.

Table 5
Revised CAINS items.

Experiential
1. Family: motivation
2. Romantic: motivation
3. Friendships: motivation
4. Social: pleasure frequency
5. Social: expected pleasure (P)
6. Vocational: motivation
7. Vocational: pleasure frequency (P)
8. Vocational: expected pleasure (P)
9. Recreational: motivation
10. Recreational: pleasure frequency
11. Recreational: expected pleasure (P)
Expressive
12. Facial expression
13. Vocal expression
14. Vocal prosody (P)
15. Gestural expression
16. Quantity of speech

Notes: P = provisional item.

In the final phase of the CANSAS project, the reliability and validity of the revised CAINS will be evaluated in another large sample. Test–retest reliability will be assessed to ensure that the scale shows sufficient stability for use as an endpoint in clinical trials. The convergent and discriminant validity of the CAINS will also be comprehensively evaluated. For example, independent ratings on the widely-used Scale for the Assessment of Negative Symptoms (Andreasen, 1983) will be conducted by raters who are blind to CAINS ratings. Participants will also be assessed on psychiatric symptoms, functional capacity, functional attainment, neurocognition, and alternative measures of emotion and motivation. The final version of the CAINS will be a major step forward in the assessment of negative symptoms, thus meeting the original charge of the consensus development conference and setting the stage for the next generation of pharmacological treatment advances. The development process of the CAINS is unique with respect to the sample size, a priori empirically driven approach, and comprehensive assessment of psychometric properties and validity.

This approach will prove fruitful for attaining the ultimate goal, which is to stimulate novel pharmacological and psychosocial treatments and new research into the underlying causes of negative symptoms. Efforts to develop neuroscience based accounts of schizophrenia (e.g., Juckel et al., 2006; Gur et al., 2007; Ochsner, 2008; Barch and Dowd, 2010) integrating psychological, neurobehavioral, neuroimaging and genomic data hinge on the integrity and reliability of clinical phenotypic data. The CAINS can contribute to treatment by providing an outcome measure for interventions aimed at ameliorating impaired emotion processing and social cognition in schizophrenia (Green et al., 2008; Carter et al., 2009). Functional neuroimaging studies examining the underlying neural circuitry of motivated behavior and affective processes would be complemented by the CAINS, which can help establish associations between activation abnormalities and deficits in specific symptom domains.

Role of funding sources

This work was supported by the National Institute of Mental Health (K02-MH079231 and R01-MH082839 to JJB; R01-MH082890 to AMK; R01-MH082782 to WPH; and R01-MH082783 to REG). Funding sources had no role in study design or in the collection, analysis and interpretation of data; or in the writing of this report.

Contributors

Drs. Blanchard, Horan, Kring, and Gur contributed to the study design, data collection, data analysis, interpretation of data analysis, and writing of this report. Dr. Reise contributed to the data analysis, interpretation of data analysis, and writing of this report.

Conflict of interest

None.

Acknowledgments

The authors would like to thank and acknowledge the following people who contributed to the data collection and environment for the study: Janelle Caponigro, Erin Moran, Timothy Campellone, and Ori Ellis at UC Berkeley; Mary March and Elizabeth Hanson at University of Pennsylvania, Medical Center; Shelly Crosby, Heidi Kupppinger, Christen Waldon, Mark McGee, Cory Tripp, Stephen Marder, and Michael F. Green at University of California, Los Angeles; Melanie Bennett, Angie Grodack, and Courtney Forbes from University of Maryland.

References

- Addington, D., Addington, J., Schissel, B., 1990. A depression rating scale for schizophrenics. *Schizophrenia Research* 3, 247–251.
- Addington, D., Addington, J., Atkinson, M., 1996. A psychometric comparison of the Calgary depression scale for schizophrenia and the Hamilton depression rating scale. *Schizophrenia Research* 19, 205–212.
- Aldenderfer, R.K., Blashfield, R.K., 1985. *Cluster Analysis*. Sage Publications, Los Angeles.
- Andreasen, N.C., 1983. *The Scale for the Assessment of Negative Symptoms (SANS)*. The University of Iowa, Iowa City, Iowa.
- Barch, D.M., Dowd, E.C., 2010. Goal representations and motivational drive in schizophrenia: the role of prefrontal–striatal interactions. *Schizophrenia Bulletin* 36, 919–934.
- Blanchard, J.J., Cohen, A.S., 2006. The structure of negative symptoms within schizophrenia: implications for assessment. *Schizophrenia Bulletin* 32, 238–245.
- Blanchard, J.J., Kring, A.M., Horan, W.P., Gur, R.E., 2011. Toward the next generation of negative symptom assessments: the collaboration for the advancement of negative symptom assessment in schizophrenia. *Schizophrenia Bulletin* 37, 291–299.
- Carter, C.S., Barch, D.M., Gur, R., Pinkham, A., Ochsner, K., 2009. CNTRICS final task selection: social cognitive and affective neuroscience-based measures. *Schizophrenia Bulletin* 35, 153–162.
- Cohen, A.S., Minor, K.S., 2008. Emotional experience in patients with schizophrenia revisited: meta-analysis of laboratory studies. *Schizophrenia Bulletin* 36, 143–150.
- Embretson, S., Reise, S.P., 2000. *Psychometric Methods: Item Response Theory for Psychologists*. Erlbaum, Mahwah, NJ.
- First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., 1996. *Structured Clinical Interview for DSM-IV Axis I Disorders Patient Edition*. New York, Biometrics Research.
- Forbes, C., Blanchard, J.J., Bennett, M., Horan, W.P., Kring, A., Gur, R., 2010. Initial development and preliminary validation of a new negative symptom measure: the Clinical Assessment Interview for Negative Symptoms (CAINS). *Schizophrenia Research* 124, 36–42.
- Goldberg, L.R., 2002. Doing it all Bass-Ackwards: the development of hierarchical factor structures from the top down. *Journal of Research in Personality* 40.
- Green, M.F., Penn, D.L., Bental, R., Carpenter, W.T., Gaebel, W., Gur, R.C., Kring, A.M., Park, S., Silverstein, S.M., Heinssen, R., 2008. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophrenia Bulletin* 34, 1211–1220.
- Gur, R.E., Loughhead, J., Kohler, C.G., Elliott, M.A., Lesko, K., Ruparel, K., Wolf, D.H., Bilker, W.B., Gur, R.C., 2007. Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. *Archives of General Psychiatry* 64, 1356–1366.
- Horan, W.P., Blanchard, J.J., Kring, A.M., 2006. Anhedonia in schizophrenia: a review of assessment strategies. *Schizophrenia Bulletin* 32, 359–373.
- Juckel, G., Schlagenhauf, F., Koslowski, M., Wustenberg, T., Villringer, A., Knutson, B., Wrase, J., Heinz, A., 2006. Dysfunction of ventral striatal reward prediction in schizophrenia. *NeuroImage* 29, 409–416.
- Kimhy, D., Yale, S., Goetz, R.R., McFarr, L., Malaspina, D., 2006. The factorial structure of the schedule for the deficit syndrome in schizophrenia. *Schizophrenia Bulletin* 32, 274–278.
- Kirkpatrick, B., Fenton, W., Carpenter, W.T., Marder, S.R., 2006. The NIMH-MATRICES consensus statement on negative symptoms. *Schizophrenia Bulletin* 32, 296–303.
- Kopelowicz, A., Ventura, J., Liberman, R.P., Mintz, J., 2008. Consistency of brief psychiatric rating scale factor structure across a broad spectrum of schizophrenia patients. *Psychopathology* 41, 77–84.
- Kring, A.M., Moran, E.K., 2008. Emotional response deficits in schizophrenia: insights from affective science. *Schizophrenia Bulletin* 34, 819–834.
- Laughren, T., Levin, R., 2006. Food and Drug Administration perspective on negative symptoms in schizophrenia as a target for a drug treatment claim. *Schizophrenia Bulletin* 32, 220–222.
- Leucht, S., Pitschel-Walz, G., Abraham, D., Kissling, W., 1999. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophrenia Research* 35, 51–68.
- Messinger, J.W., Tremereau, F., Antonius, D., Mendelsohn, E., Prudent, V., Stanfor, A.D., Malaspina, D., 2011. Avolition and expressive deficits capture negative symptom phenomenology: implications for DSM-5 and schizophrenia research. *Clinical Psychology Review* 31, 161–168.
- Montgomery, S.A., Zwieter-Boot, B., 2006. ECNP consensus meeting. Negative, depressive, and cognitive symptoms of schizophrenia. Nice, March 2004. *European Neuropsychopharmacology* 17, 70–77.
- Ochsner, K.N., 2008. The social-emotional processing stream: five core constructs and their translational potential for schizophrenia and beyond. *Biological Psychiatry* 64, 48–61.
- Overall, J.E., Gorham, D.R., 1962. The brief psychiatric rating scale. *Psychological Reports* 10, 799–812.
- Reise, S.P., Ainsworth, A.T., Haviland, M.G., 2005. Item response theory: fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science* 14, 95–101.
- Samejima, F., 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement* 34, 100–114.
- Sijtsma, K., Molenarr, I.W., 2002. *Introduction to Non-parametric Item Response Theory*. Sage, Thousand Oaks, CA.
- Thissen, D., 2003. *MULTILOG 7.03*. Scientific Software, Chicago.
- Wechsler, D., 2001. *Wechsler Test of Adult Reading*. Psychological Corporation, London.