Schizophrenia Bulletin vol. 49 no. 5 pp. 1205–1216, 2023
https://doi.org/10.1093/schbul/sbad038
Advance Access publication April 25, 2023

Development and Validation of the Negative Symptom Inventory-Psychosis Risk

Gregory P. Strauss[*,1], Elaine F. Walker[2], Andrea Pelletier-Baldelli[3], Nathan T. Carter[4], Lauren M. Ellman[5], Jason Schiffman[6], Lauren Luther[1], Sydney H. James[1], Alysia M. Berglund[1], Tina Gupta[7], Ivanka Ristanovic[7], and Vijay A. Mittal[7]

[1]Department of Psychology, University of Georgia, Athens, GA, USA; [2]Department of Psychology, Emory University, Atlanta, GA, USA; [3]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; [4]Department of Psychology, Michigan State University, East Lansing, MI, USA; [5]Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, USA; [6]Department of Psychological Science, University of California- Irvine, Irvine, CA, USA; [7]Department of Psychology, Northwestern University, Evanston, IL, USA

[*]To whom correspondence should be addressed; University of Georgia, Department of Psychology, 125 Baldwin St., Athens, GA 30602, USA; tel 1-706-542-0307, fax: 1-706-542-3275, e-mail: gstrauss@uga.edu

***Background and Hypotheses*: Early identification and prevention of psychosis is limited by the availability of tools designed to assess negative symptoms in those at clinical high-risk for psychosis (CHR). To address this critical need, a multi-site study was established to develop and validate a clinical rating scale designed specifically for individuals at CHR: The Negative Symptom Inventory-Psychosis Risk (NSI-PR). *Study Design*: The measure was developed according to guidelines recommended by the NIMH Consensus Conference on Negative Symptoms using a transparent, iterative, and data-driven process. A 16-item version of the NSI-PR was designed to have an overly inclusive set of items and lengthier interview to support the ultimate intention of creating a new briefer measure. Psychometric properties of the 16-item NSI-PR were evaluated in a sample of 218 CHR participants. *Study Results*: Item-level analyses indicated that men had higher scores than women. Reliability analyses supported internal consistency, inter-rater agreement, and temporal stability. Associations with measures of negative symptoms and functioning supported convergent validity. Small correlations with positive, disorganized, and general symptoms supported discriminant validity. Structural analyses indicated a 5-factor structure (anhedonia, avolition, asociality, alogia, and blunted affect). Item response theory identified items for removal and indicated that the anchor range could be reduced. Factor loadings, item-level correlations, item-total correlations, and skew further supported removal of certain items. *Conclusions*: These findings support the psychometric properties of the NSI-PR and guided the creation of a new 11-item NSI-PR that will be validated in the next phase of this multi-site scale development project.**

Keywords: prodrome/rating scale/clinical interview/clinical trial/early identification/prevention

Introduction

Negative symptoms are the most significant predictor of functional disability in psychotic disorders;[1,2] however, attempts to assess, understand, and treat this symptom domain have been ineffective.[3,4] To address this critical unmet need, national institute of mental health (NIMH) sponsored a consensus development conference in 2005.[5] A critical outcome of this meeting was the creation of 2 next-generation clinical rating scales for adults with psychotic disorders, which assess 5 core domains (anhedonia, avolition, asociality, blunted affect, and alogia): the Brief Negative Symptom Scale[6] (BNSS) and Clinical Assessment Interview for Negative Symptoms[7] (CAINS). These instruments have provided an improved assessment approach that has facilitated the development of targeted treatments for negative symptoms in adults with schizophrenia.[8]

However, the consensus conference did not discuss the development of assessments specific to those at clinical high risk (CHR) for developing psychosis (ie, adolescents meeting criteria for a psychosis-risk syndrome). Improved assessment of negative symptoms may be crucial for enhancing early identification and prevention efforts for several reasons. For example, among CHR participants, negative symptoms are: (1) highly prevalent (eg, present in 82% of CHR cases in the north atlantic prodromal longitudinal study); (2) a strong predictor of conversion; (3) associated with poor social and role functioning; (4) a factor that brings individuals into initial contact with the treatment system, often years before the emergence

of attenuated positive symptoms.[9–12] Given the aforementioned evidence, accurate detection of negative symptoms in CHR individuals is paramount because it will improve our ability to accurately chart mental illness trajectories and determine when, where, and how to intervene more effectively to prevent psychotic disorders.

Unfortunately, instruments used to assess negative symptoms in the CHR population are not ideal for these purposes. The Structured Interview for Psychosis-risk Syndromes[13] (SIPS) and Comprehensive Assessment of At Risk Mental States[14] are the instruments most widely used to assess the presence of psychosis-risk syndromes and rate symptom severity in the CHR population. Although these measures are well validated as measures of psychosis risk, widely used, and have been central to advances in the early identification and prevention of psychosis, there are several conceptual, methodological, and psychometric limitations associated with their negative symptom subscales (for a review see[15]). Attempts to adapt the CAINS[16] and BNSS,[17] to CHR participants have also been suboptimal, potentially due to the process of adaptation itself (ie, only probes were modified and anchors were not adjusted from the range designed for schizophrenia, leading to mean item scores for CHR participants that generally approached the scale's floor and highly positively skewed item and subscale scores). In addition, subscales on the adapted CAINS and BNSS showed modest to weak convergent and discriminant validity (eg, higher correlations with general symptoms than negative symptoms).[16,17] Accordingly, the European Psychiatric Association's[8] and others[15] have advocated for new negative symptom scales designed specifically for CHR individuals.

The current study extended the scale development aims of the NIMH consensus development conference to the CHR population and created a new measure, the Negative Symptom Inventory for Psychosis Risk (NSI-PR). This study includes the first steps of a multi-stage, iterative scale development and validation process, with the overall goal of arriving at a final version of the NSI-PR by using the type of transparent and data-driven approach that was recommended by the NIMH consensus conference. We review the process for developing the measure and report the results of a psychometric evaluation of a 16-item version of the NSI-PR, which was created to have an overly inclusive set of items and lengthier interview that will be trimmed based on psychometric considerations. The end goal of this process is to create a briefer final version of the scale with strong psychometric properties that is appropriate for wide-spread dissemination and implementation in CHR research and clinical work.

Method

Participants

Two hundred and eighteen CHR participants were recruited from 5 sites: (1) Georgia Psychiatric Risk Evaluation Program in Athens, GA ($n = 67$); (2) Northwestern University Adolescent Development and Preventative Treatment research program in Evanston, IL ($n = 66$); (3) Mental Health and Development Program at Emory University in Atlanta, GA ($n = 11$); (4) Youth FIRST research program at the University of Maryland, Baltimore County ($n = 27$), and (5) Ellman Lab at Temple University ($n = 50$). See table 1 for participant demographics. CHR participants were eligible if they met the criteria for a psychosis-risk syndrome on the (SIPS).[13] No participants met lifetime criteria for a DSM-5 full psychotic disorder based on the structured clinical interview for the diagnostic and statistical manual of mental disorders-5 (SCID-5).[18]

Procedures

All participants provided written informed consent for studies approved by institutional review boards and received monetary compensation for participation. Participants completed clinical interviews conducted by the PIs or raters trained using gold-standard training videos. In addition to the SCID-5, SIPS, and NSI-PR, participants were rated on the Global Functioning Scale: Social[19] (GFS:S), Global Functioning Scale: Role[20] (GFS:R), and Global Assessment of Functioning (GAF).

NSI-PR Scale Development

In developing the initial version of the NSI-PR, we took a broad, overly inclusive approach to developing items. First, items were created to specifically cover all 5 negative symptom domains. Second, anchors and probes were

Table 1. Participant Demographic and Clinical Characteristics

	Mean (SD)
Age	21.1 (3.2)
Education	13.8 (2.0)
% Female	67.4%
% Hispanic	12.8%
Race	
Black	19.0%
Asian	14.0%
Biracial	6.3%
Central/South American	4.1%
Native American	1.4%
White	53.4%
SIPS positive total	11.0 (3.7)
SIPS negative total	7.2 (5.02)
SIPS general total	7.5 (4.3)

Note: This included Attenuated Positive Symptoms Syndrome ($n = 213$), Genetic Risk and Deterioration Syndrome (GRD; n = 3), and multiple syndromes: APSS + BIPS (n = 1), APSS + BIPS + GRD (n = 1). Of these participants, 121 CHR participants met criteria for progression, 91 for persistence, 2 for partial remission, 1 for full remission, and 3 were unknown (all GRD).

worded to target youth/young adults. Third, asociality items were included for social media/texting behavior. Fourth, both the avolition and asociality domains include separate items for internal experience and behavior, as dissociations between these variables may be prognostically important and useful for identifying secondary negative symptoms. Fifth, based on our preliminary data indicating that a lack of distress in the face of negative symptoms is indicative of psychosis risk and a poorer prognosis,[21] a lack of transitionary distress item was created. A single lack of transitory distress item was added following BNSS scale procedures and BNSS psychometric evaluations where a single transitory distress item met standard psychometric evaluation criteria.[6,22] Finally, based on contemporary affective science,[23,24] anhedonia items were added that separate anticipatory and past-week pleasure in relation to recreational, role, social, and physical activities to capture multiple aspects of hedonic response that are impaired in CHR.[25–28] These procedures resulted in a 16-item NSI-PR version. Items are rated based on a semi-structured interview on a 0 (absent) to 6 (extremely severe) scale.

Analyses

Several analytic steps were conducted to identify how the NSI-PR should be revised. Since the initial NSI-PR was created to contain an oversampling of items from the 5 consensus conference negative symptom domains, analyses focused on item selection, modification, and retention. First, analyses were conducted to evaluate the functioning of individual items. Item means, standard deviations, skewness, kurtosis, and frequency distributions were calculated to evaluate floor and ceiling effects. Inter-item and item-total correlations were calculated to determine the degree to which individual items were associated with each other and the total score, respectively. Sex differences were evaluated using MANOVA.

Reliability was evaluated using internal consistency, inter-rater reliability, and temporal stability. Within and between site inter-rater reliability was examined with intra-class correlation coefficients (ICC). Temporal stability was examined via correlations in a subset of participants retested at 1 year.

Convergent validity was examined at the item, domain, and total level with the SIPS negative symptom dimension, GFS:S, GFS:R, and GAF. Discriminant validity was evaluated in relation to (low) correlations with the SIPS positive, disorganized, general symptom dimensions.

Construct validity was assessed using 2 sets of analyses. First, we utilized confirmatory multidimensional item response theory modeling (MIRT) based on the graded response model[29] (GRM). To account for the number of dimensions,[30] GRM parameters were estimated using diagonally weighted least squares in the "lavaan" package,[31] and estimates were converted from the "item factor analysis" to the "item response theory" scaling using simple algebraic formulae.[32] Second, IRT was used to identify (a) items that could be deleted without considerable loss of content coverage and reliability at higher score levels and (b) to examine whether all response options in the 6-option response format were discriminating between participants with higher vs lower negative symptoms through examination of the option characteristic curve (OCC), which relates response probabilities to symptom levels for each option for a given item.

Factor structure of the NSI-PR was also tested using the GRM in MIRT. We tested three alternative confirmatory models: (a) a 1-factor model; (b) a 2-factor model with Motivation and Pleasure and Diminished Expression factors; (c) a 5-factor model based on the consensus domains.[29] All models were estimated using quasi-Monte Carlo expectation-maximization (QMCEM) due to their better performance in high-dimensional models. We also ran hierarchical confirmatory factor analysis (CFA) to determine whether the measure showed a hierarchical structure. Fit indices for each model were evaluated using the multi-index approach:[33] Comparative Fix Index (CFI) > 0.90; Root Mean Square Error of Approximation (RMSEA) values < 0.08, and Tucker-Lewis Index (TLI) values > 0.90 are acceptable. Models with fit indices at least approximating this criteria were deemed important to further evaluate in the final NSI-PR scale validation.

Collectively, these analyses guided decisions for creating a final, shorter NSI-PR. Items were considered for removal or revision if they demonstrated: (1) floor/ceiling effects, (2) high skew (> ±1.0), (3) low inter-item or item-total correlations ($r < 0.30$), (4) redundancy with other items as indicated by high item-level correlations ($r > 0.55$), (5) low inter-rater reliability within- and between-sites (< 0.70), (6) low factor loadings (< 0.4) or high cross-loadings, (7) poor convergent validity with the SIPS negative subscale and GFS; (8) poor discriminant validity with SIPS positive, disorganized, and general subscales; (9) poor ability to distinguish participants with scores 1 SD or greater above the mean NSI-PR total score. Participant and interviewer feedback was also incorporated.

Results

Item-Level Descriptive Statistics

Table 2 reports means, standard deviation, skewness, and kurtosis for all 16 items. Most items were not highly skewed. However, some skew > 1.0 was observed for items 5 (asociality behavior), 6 (asociality inner experience), 9 (anhedonia past week pleasure), 11 (anhedonia affective forecasting), 14 (blunted vocal affect), and 16 (alogia). Mean scores were low (< 1.0) for 2 items: 11 (affective forecasting) and 16 (alogia). Avolition behavior role, avolition internal experience role, asociality inner

experience, and alogia items did not receive the full range of scores, although all others achieved the full range of ratings.

Sex Differences

The overall omnibus effect was significant, $F(16,199) = 1.93$, $P = .02$ (partial eta squared = 0.13), such that males scored higher than females. Differences were primarily driven by the blunted affect items, but males also had greater asociality via social/electronic media use behavior (table 3). Females had greater severity than males on avolition behavior for recreational activities.

Inter-Item Correlations

To examine redundancy among items, inter-item correlations were calculated among all 16 items (table 4). The avolition role and recreation internal experience and behavior items were highly correlated (as expected), as were the in-person and social media/electronic asociality internal experience and behavior items. Lack of transitionary distress had low correlations with all items. The 3 blunted affect items all had high inner correlations, with the blunted facial and vocal affect being the highest. The lone alogia item was moderately correlated, but not redundant, with the 3 blunted affect items.

Inter-rater Agreement

Inter-rater agreement was evaluated in relation to 8 video-recorded interviews that were rated by 22 raters across the 5 sites. For the full scale, inter-rater agreement across all sites was 0.97. Item-level inter-rater reliability ranged from 0.90 (asociality internal experience social media) to 1.00 (anhedonia affective forecasting) (table 5).

Inter-rater reliability calculated within each site was 0.90 on average: University of Georgia ($n = 5$) = 0.92; Northwestern University ($n = 5$) = 0.97; Emory University ($n = 6$) = 0.89; University of Maryland, Baltimore County ($n = 3$) = 0.80; Temple University ($n = 3$) = 0.93.

Internal Consistency

Cronbach's α was = 0.85 for the total NSI-PR score and avolition = 0.78; asociality = 0.79; anhedonia = 0.58; blunted affect = 0.88. Alpha-if-item deleted analyses suggested that removing lack of transitionary distress would improve internal consistency, whereas removing other items would worsen it.

Temporal Stability

Stability across 1 year was high ($r = 0.77$) for total scores in a subset of participants ($n = 30$) for whom 1-year longitudinal data were available. Domain score stability was: avolition: $r = 0.43$, $P < 0.03$; anhedonia: $r = 0.49$, $P < 0.01$; asociality: $r = 0.66$, $P < .001$; blunted affect: $r = 0.59$, $P < .001$; alogia: $r = 0.30$, $P = .13$.

Convergent Validity

The NSI-PR total score demonstrated moderate convergent validity with the SIPS negative dimension, GFS-Social, GFS-Role, and GAF (table 6). At the domain level, moderate convergent validity was generally observed with the SIPS negative dimension (table 6) and individual SIPS items (supplementary table 1). Item-level correlations (supplementary table 2) of similar constructs between the NSI-PR and SIPS generally demonstrated moderate convergence. Notably, the lack of transitionary

Table 2. Item-level Descriptive Statistics

Item	Mean	Std. Dev	Skewness	Kurtosis	Item-Total Correlation
1 Avolition behavior role	1.78	1.43	0.50	−0.65	0.52***
2 Avolition internal experience role	1.39	1.17	0.35	−0.90	0.47***
3 Avolition behavior recreation	1.80	1.58	0.72	−0.13	0.47***
4 Avolition internal experience recreation	2.02	1.50	0.32	−0.72	0.47***
5 Asociality behavior	1.45	1.34	**1.04**	0.47	0.52***
6 Asociality internal experience	1.22	1.15	**1.06**	0.90	0.62***
7 Asociality behavior social media	1.56	1.39	0.77	0.29	0.46***
8 Asociality internal experience social media	2.11	1.48	0.36	−0.66	0.54***
9 Anhedonia past week intensity	1.39	1.49	**1.29**	1.24	0.48***
10 Anhedonia past week frequency	2.69	1.64	−0.36	−1.00	**0.37*** **
11 Anhedonia affective forecasting	0.69	1.12	**2.26**	6.02	0.45***
12 Lack of transitionary distress	2.39	1.81	0.52	−0.46	**0.22*** **
13 Blunted facial affect	1.41	1.40	0.70	−0.49	0.57***
14 Blunted vocal affect	1.23	1.48	**1.05**	0.16	0.56***
15 Gestural expression	1.61	1.67	0.71	−0.70	0.50***
16 Alogia	0.74	1.11	**1.48**	1.54	0.51***

Note: Bold = exceeded a priori criteria for skew or criteria for low item-total correlations.
*** = P < .001.

Table 3. Sex Differences

	Female		Male				
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	***F***	***P*-value**	**Par Eta-Sq**
1 Avolition behavior role	1.80	1.46	1.72	1.39	0.16	.69	0
2 Avolition internal experience role	1.38	1.15	1.44	1.23	0.13	.72	0
3 Avolition behavior recreation	**1.99**	**1.63**	**1.43**	**1.42**	**6.12**	**.01**	**0.03**
4 Avolition internal experience recreation	2.15	1.50	1.76	1.49	3.08	.08	0.01
5 Asociality behavior	1.45	1.39	1.49	1.25	0.04	.84	0
6 Asociality internal experience	1.17	1.19	1.31	1.10	0.68	.41	0
7 Asociality behavior social media	**1.37**	**1.34**	**1.93**	**1.43**	**7.69**	**.01**	**0.04**
8 Asociality internal experience social media	1.99	1.48	2.34	1.46	2.56	.11	0.01
9 Anhedonia past week intensity	1.34	1.46	1.41	1.48	0.12	.73	0
10 Anhedonia past week frequency	2.74	1.66	2.56	1.62	0.58	.45	0
11 Anhedonia affective forecasting	0.60	1.15	0.88	1.04	2.95	.09	0.01
12 Lack of transitionary distress	2.29	1.83	2.57	1.72	1.16	.28	0.01
13 Blunted facial affect	**1.23**	**1.31**	**1.81**	**1.51**	**8.26**	**.01**	**0.04**
14 Blunted vocal affect	**1.09**	**1.46**	**1.54**	**1.48**	**4.53**	**.04**	**0.02**
15 Gestural expression	**1.41**	**1.65**	**2.03**	**1.65**	**6.51**	**.01**	**0.03**
16 Alogia	0.73	1.10	0.78	1.16	0.09	.76	0

Note: Bold = items with significant sex differences.

distress item did not show good convergent validity at the domain or item level, potentially suggesting that it should be considered for removal.

Discriminant Validity

The NSI-PR total score, domains, and items demonstrated good discriminant validity via low correlations with positive, disorganized, and general symptoms (table 7 and supplementary table 3).

Item Response Theory

IRT was used to identify: (a) items that could be deleted without considerable loss of content coverage, and reliability at higher levels of scores, and (b) to examine whether all options in the 6-option response format are discriminating between patients with higher vs lower levels of the traits.

With regard to (a), we found that items 4 (avolition internal experience recreation), 10 (anhedonia past week frequency), and 12 (lack of transitionary distress) were the best candidates for deletion. These items primarily impart reliability for those between −0.15 and + 0.38 away from the mean, and thus are not likely to be helpful in distinguishing between those in the more severe range. A second set that was identified in addition to these 3, which included item 3 avolition behavior recreation), 8 (asociality internal experience social media), 13 (blunted facial affect), and 14 (blunted vocal affect). These items primarily help distinguish between those between + 0.54 and + 0.75 away from the mean. All other items are necessary to distinguish those falling 1 SD or greater above the mean, and thus should be kept to help distinguish between those falling within the more and less severe symptom ranges.

With regard to (b), we found that in general, the NSI-PR would function equivalently if options 2 and/or 3 were discarded. Figure 1 shows an example of this phenomenon which was found in most NSI-PR items, especially those identified in (a) as items that should be retained. As can be seen, the OCC relates response probabilities to trait levels for each option for a given item. Importantly, there is very little space wherein options 2 and 3 do not overlap with other options; this is partially due to underuse of these options by raters, and indicates that they are not particularly useful in reliably discriminating between those at more moderate levels of the underlying trait.

Scale Structure

The NSI-PR was also analyzed utilizing multidimensional item response theory. Specifically, we fit the graded response model[28] (GRM), a polytomous IRT model that proposes item response are ordinal, and that items may differ in their strength of connection or "loading" to the latent variable(s). We tested three alternative confirmatory models: (a) a 1-factor model for all 15 items (items 1–11 and 13–16); (b) a 2-factor model that divides the measure into Motivation and Pleasure (items 1–11) and Diminished Expression (13–16) factors; (c) a 5-factor model that divides the measure into avolition (items 1–4), asociality (items 5–8), anhedonia (items 9–11), blunted affect (items 13–15), and alogia (item 16) factors. All models were estimated using QMCEM due to their better performance in high-dimensional models.

As shown in table 8, the 1- and 2-factor models showed poor model-data fit, suggesting aggregation across all items or division by Motivation and Pleasure and Diminished Expression may be misleading in practice.

Table 4. Inter-item Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00	0.62***	0.38***	0.45***	0.31***	0.33***	0.23**	0.23**	0.29***	0.25***	.32***	.11	.20**	.22**	.22**	.27***
2	0.62***	1.00	0.26***	0.38***	0.24***	0.35***	0.28***	0.23***	0.24***	0.22**	.29***	.14*	.16*	.14*	.25***	0.19**
3	0.38***	0.26***	1.00	0.70***	0.33***	0.31***	0.12	0.29***	0.24***	0.25***	.27***	.14*	.17*	.19**	.08	.24***
4	0.45***	0.38***	0.70***	1.00	0.24***	0.26***	0.13	0.25***	0.19**	0.26***	.22**	.21**	.18**	.18**	.07	.22**
5	0.31***	0.24***	0.33***	0.24***	1.00	0.65***	0.41***	0.39***	0.25***	0.15*	.31***	.08	.29***	.25***	.32***	.28***
6	0.33***	0.35***	0.31***	0.26***	0.65***	1.00	0.46***	0.51***	0.31***	0.28***	.34***	.15*	.35***	.33***	.33***	.30***
7	0.23**	0.28***	0.12	0.13	0.41***	0.46***	1.00	0.53***	0.28***	0.10	.28***	0.08	.31***	.27***	.30***	.19**
8	0.23**	0.23***	0.29***	0.25***	0.39***	0.51***	0.53***	1.00	0.25***	0.29***	.31***	.26***	.30***	.30***	.24***	0.14*
9	0.29***	0.24***	0.24***	0.19**	0.25***	0.31***	0.28***	0.25***	1.00	0.36***	.43***	.09	.23**	.27***	.30***	.32***
10	0.25***	0.22**	0.25***	0.26***	0.15*	0.28***	0.10	0.29***	0.36***	1.00	.20**	.18**	.14*	.21**	.11	0.16*
11	0.32***	0.29***	0.27***	0.22**	0.31***	0.34***	0.28***	0.31***	0.43***	0.20**	1.00	.09	.22**	.17*	.15*	.29***
12	0.11	0.14*	0.14*	0.21**	0.08	0.15*	0.08	0.26***	0.09	0.18**	0.09	1.00	.18**	.11	.07	.01
13	0.20**	0.16*	0.17*	0.18**	0.29***	0.35***	0.31***	0.30***	0.23**	0.14*	.22**	.18**	1.00	.82***	.67***	.55***
14	0.22**	0.14*	0.19**	0.18**	0.25***	0.33***	0.27***	0.30***	0.27***	0.21**	.17*	.11	.82***	1.00	.67***	.58***
15	0.22**	0.25***	0.08	0.07	0.32***	0.33***	0.30***	0.24***	0.30***	0.11	.15*	.07	.67***	.67***	1.00	.56***
16	0.27***	0.19**	0.24***	0.22**	0.28***	0.30***	0.19**	0.14*	0.32***	0.16*	.29***	.01	.55***	.58***	.56***	1.00

Note: Bold and underlined = items exceeding the $r = \pm 0.55$ cutoff. *$P < .05$, ** $P < .01$, ***$P < .001$.

Table 5. Item-level Inter-rater Reliability

	ICC
1 Avolition behavior role	0.97
2 Avolition internal experience role	0.98
3 Avolition behavior recreation	0.98
4 Avolition internal experience recreation	0.96
5 Asociality behavior	0.99
6 Asociality internal experience	0.95
7 Asociality behavior social media	0.97
8 Asociality internal experience social media	0.90
9 Anhedonia past week intensity	0.99
10 Anhedonia past week frequency	0.99
11 Anhedonia affective forecasting	1.00
12 Lack of transitional distress	0.98
13 Blunted facial affect	0.97
14 Blunted vocal affect	0.97
15 Gestural expression	0.97
16 Alogia	0.97

The 5-factor model on the other hand showed very good fit, with RMSEA = 0.04, and TLI and CFI > 0.95. Table 9 shows the parameter estimates from this model. In the multidimensional GRM, there is a set of item loading (or "discrimination") estimates for each latent trait estimated, denoted *a*, but only one set of threshold parameters (1 less than the number of response options), denoted *b*. Thresholds indicate the level of the latent trait at which an option becomes more likely to be chosen than the one below; for example, b_1 is the level of the trait at which one is equally likely to choose option 2 over option 1. As can be seen, item loadings were high, and threshold parameters were largely reasonable, ranging from −1.78 to 5.74; spaces where no threshold is given indicates that there was not enough response data for those options to estimate the threshold.

Marginal reliability (ie, IRT-based reliability) ranged from .75 to .88, with the exception of the single-item factor, which showed a .59 reliability; this is a relatively high reliability estimate for a single item. Figure 2 shows the test information functions for all subscales of the NSI-PR. Notably, for all subscales, the IRT information—the inverse of the IRT standard error of measurement and the basis for marginal reliability—suggests reliability is highest at higher levels of the measured variable. For example, although Alogia had an overall reliability of .59, this reliability level applies to the *average* respondent, and therefore reliability for respondents above the mean would be higher. Of course, this supports the practical utility of the measure in that we are most concerned with making distinctions among those who receive *high* rather than *low* scores. Notably, it is important not to compare the information values on the *y*-axis from different subscales, as their levels are in part determined by the number of items.

Finally, we wanted to determine the extent to which the measure showed a hierarchical structure. The IRT

Table 6. Domain and Total Score Convergent Validity

	SIPS Negative	GFS Social	GFS Role	GAF Current	GAF Last Year
NSI-PR avolition	0.54***	−0.42***	−0.51***	−0.43***	−0.33***
NSI-PR asociality	0.56***	−0.59***	−0.25**	−0.35***	−0.25**
NSI-PR anhedonia	0.36***	−0.33***	−0.24**	−0.25**	−0.06
NSI-PR transitionary distress	−0.03	−0.09	−0.03	−0.02	0.00
NSI-PR blunted affect	0.38***	−0.26**	−0.19*	0.06	0.11
NSI-PR alogia	0.36***	−0.25**	−0.21*	−0.02	0.04
NSI-PR total	0.62***	−0.53***	−0.40***	−0.30***	−0.17*

Note: GFS, Global Functioning Scale; GAF, Global Assessment of Functioning; NSI-PR, Negative Symptom Inventory-Psychosis Risk. *P < .05, ** P < .01, ***P < .001.

Table 7. Domain Level Discriminant Validity

	SIPS Positive	SIPS Disorganized	SIPS General
NSI-PR avolition	0.09	0.25**	0.34***
NSI-PR asociality	0.20**	0.27***	0.21*
NSI-PR anhedonia	−0.06	0.10	0.15
NSI-PR transitionary distress	−0.13	−0.05	−0.14
NSI-PR blunted affect	0.08	0.04	0.02
NSI-PR alogia	0.04	−0.05	0.15
NSI-PR total	0.10	0.19*	0.22**

Note: NSI-PR, Negative Symptom Inventory-Psychosis Risk; SIPS, Structured Interview for Prodromal Syndromes.
*P < .05, ** P < .01, ***P < .001.

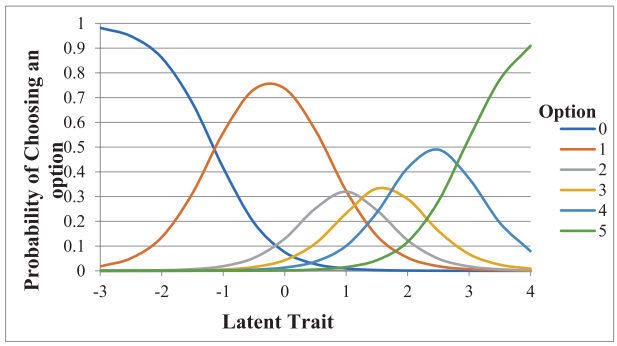

Fig. 1. Option response functions for item 5 as an example of options 2 and 3 not discriminating between respondents. Plot extends to +4 SD due to the emphasis on severity in the NSI-PR.

Table 8. Model-data Fit of the Graded Response Model to NSI-PR Items by Factor Structure

					95% CI of RMSEA			
Model	M_2	*df*	*P*	RMSEA	Low	High	TLI	CFI
One-Factor	111.99	21	< .001	0.142	0.116	0.168	0.629	0.784
Two-Factor	114.63	21	< .001	0.144	0.119	0.170	0.618	0.777
Five-Factor	30.39	22	.109	0.042	0.000	0.075	0.967	0.980

analyses indicated a 5-factor model as the only model with acceptable fit, but we wondered whether aggregation across constructs would be justified. To do this, we turned to hierarchical CFA. First, we estimated a simple 5-factor CFA analogous to our best-fitting IRT model. This model showed marginally acceptable fit (RMSEA = 0.08, CFI = 0.90, TLI = 0.88), emphasizing the importance of IRT modeling for fitting item response data. We then fit 2 alternative hierarchical factor models: (a) a model where avolition, asociality, and anhedonia were subsumed by a higher-order Motivation and Pleasure factor, whereas blunted affect and alogia were subsumed by a higher-order Diminished Expression factor; and (b) a model with a single higher-order factor subsuming all 5 subdimensions. There was little difference in model-data fit between the 5-factor model and the model with 2 higher-order factors (supplementary table 4). The loadings of the 5 subdimensions onto the 2 higher-order factors ranged from .65 to .83, suggesting that reliable composites consistent with 2 higher-order factors are justified. This was not true for the model with a single higher-order factor, and thus an overall sum score is not suggested for use. Rather, the Motivation and Pleasure and Diminished Expression subscales should be considered the most general scores that should be calculated.

Discussion

The goal of the current study was to develop and validate a new clinical interview designed to measure negative symptoms in those at CHR based on a transparent, iterative, data-driven process.

Table 9. Parameter Estimates From the 5-Dimension Graded Response Model

Item Name	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5	b_6
1 Avolition behavior role	1.17	.	.	.	.	−1.43	−0.02	0.90	2.16	3.32	.
2 Avolition internal experience role	.96	.	.	.	.	−1.15	0.23	1.80	4.02	.	.
3 Avolition behavior recreation	1.95	.	.	.	.	−0.95	−0.03	0.72	1.68	2.33	3.02
4 Avolition internal experience recreation	2.40	.	.	.	.	−1.19	−0.27	0.39	1.39	2.37	3.41
5 Asociality behavior		1.84	.	.	.	−1.00	0.57	1.17	1.91	2.85	4.01
6 Asociality internal experience		2.70	.	.	.	−0.67	0.59	1.46	2.13	2.94	.
7 Asociality behavior social media		1.37	.	.	.	−0.91	0.12	1.26	2.36	3.27	3.97
8 Asociality internal experience social media		1.44	.	.	.	−1.74	−0.35	0.40	1.51	2.47	4.11
9 Anhedonia past week intensity		.	2.08	.	.	−0.57	0.46	1.23	1.68	2.11	2.68
10 Anhedonia past week frequency		.	1.08	.	.	−1.78	−1.08	−0.52	0.41	2.42	4.46
11 Anhedonia affective forecasting		.	1.53	.	.	0.33	1.43	2.31	2.77	3.29	3.80
13 Blunted facial affect		.	.	4.90	.	−0.36	0.16	0.73	1.30	2.00	.
14 Blunted vocal affect		.	.	5.38	.	−0.09	0.34	0.82	1.28	1.76	2.64
15 Gestural expression		.	.	2.51	.	−0.38	0.24	0.62	1.02	1.91	2.61
16 Alogia		.	.	.	1.00	0.49	1.32	2.74	3.59	5.74	.
Construct Name	F1	F2	F3	F4	F5						
F1—Avolition	0.84										
F2—Asociality	0.45	0.82									
F3—Anhedonia	0.40	0.51	0.75								
F4—Blunted affect	0.22	0.47	0.43	0.88							
F5—Alogia	0.19	0.35	0.55	0.58	0.59						

Psychometric Summary

MIRT and CFA indicated that 1- and 2-factor models produced poor fit for the data. However, 5 factor (anhedonia, avolition, asociality, alogia, and blunted affect) and hierarchical (2 second-order factors reflecting diminished expression and motivation and pleasure, and 5 first-order factors reflecting the 5 consensus domains) models fit the data well. These findings add to emerging evidence on the latent structure of negative symptoms that supports the 5-factor and hierarchical models over 1- and 2-factor models.[34–46]

Reliability analyses indicated that internal consistency was adequate, but not good for the total score and domain scores. Alpha-if-item deleted analyses indicated that removal of the lack of transitionary distress item would improve reliability. Although based on prior negative symptom scale procedures.[6,22] inclusion of a single lack of transitory distress item and divergence of this items content (including interview questions) from other NSI-PR items may have resulted in poorer item performance. Inter-rater agreement was good for the total score and all individual items. Temporal stability assessed in a subset of participants who had completed 1-year follow-up evaluations was good for the total score and domain scores. Discriminant validity was indicated by low correlations with positive, disorganized, and general symptoms. Convergent validity was supported by associations with other negative symptom measures, functional outcome, and GAF. The observed correlations were high enough to demonstrate convergent validity, but not so high as to demonstrate redundancy with the SIPS negative dimension. Although one would typically want to observe correlations > 0.80 to demonstrate good convergent validity, correlations lower than this with the SIPS negative may not be evidence of poor convergent validity because of conceptual issues inherent to the SIPS negative subscale.[15] The moderate correlations therefore support that the NSI-PR is measuring the construct differently than the SIPS, as intended. Item response theory identified multiple items for removal and indicated that the anchor range could be reduced.

Scale Revisions

Collectively, these findings suggest that although the scale is largely functioning as expected, some psychometric issues need to be addressed with a revised version, as intended. To address these issues, a revised and briefer scale is needed with modified item probes, anchors, and scaling. Based on the results of the current study, the authors created such a scale that is to be psychometrically validated in the next phase of this multi-step scale development process. The following summarizes the rationale for the key changes that were implemented to the 16-item version of the scale to derive the new, final 11-item NSI-PR (table 10 for summary):

Shortening the Scale. The interview length of the 16-item scale was intentionally longer than what we intended for the final version, averaging ~ 30 min (range ~25–45). We shortened the scale by removing items (reducing from 16 to 11), reducing the number of probes, and reducing the number of pages in the manual. Based on initial data collection we have conducted for study 2 (validation of the

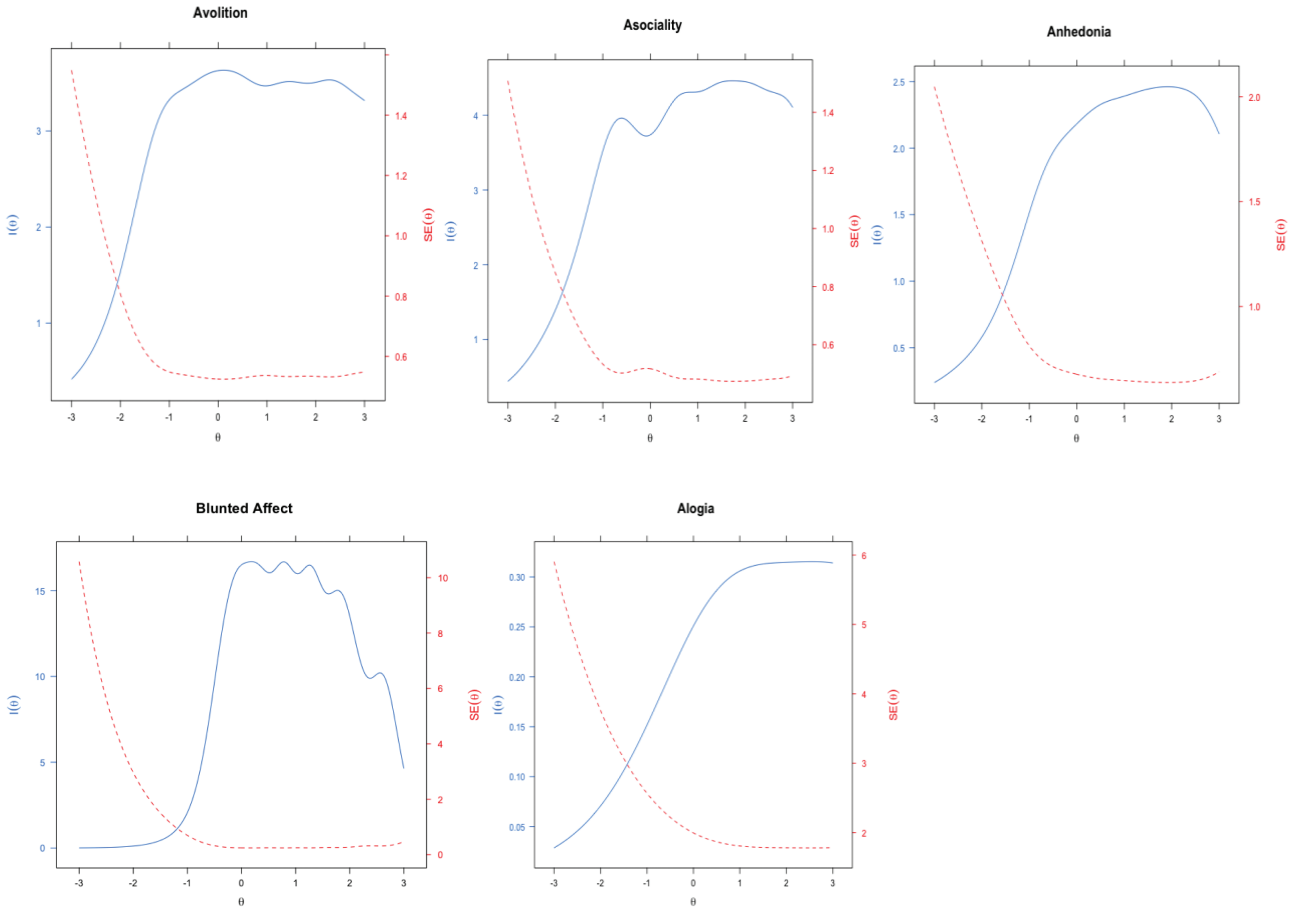

Fig. 2. Test information functions for each domain.

final 11-item scale), the estimated interview time for the final 11-item version is 15 min.

Incorporating Feedback From Interviewers and Participants. Participants found the interview tolerable, but when paired with other measures, it was a bit long. Interviewers suggested removing and combining certain items. For example, removing Lack of Transitionary Distress and combining Asociality Social Media Internal Experience with a more general Asociality Internal Experience Item.

Modifying/Trimming Items Based on Structural/Item-level Psychometrics. Psychometric analyses support removing some items based on the following criteria: (1) low factor loading communalities; (2) correlation > 0.55 with another item; (3) item-total scale correlation < 0.35; (4) skew > ±1.0. Based on these criteria the following items were identified for modification or removal.

a. *Criteria #1 Factor Loadings*: Lack of Transitionary Distress
b. *Criteria #2 High Item-Level Correlations*: Blunted Vocal Affect (note internal experience and behavior items for avolition and asociality were expected to be high)
c. *Criteria #3 Low Item Total Correlation*: Lack of Transitionary Distress
d. *Criteria #4 Skew*: Asociality Behavior, Asociality Inner-Experience, Anhedonia Affective Forecasting, Blunted Vocal Affect, Alogia

Item Removal/Modification Based on Item Response Theory. The best candidates for removal were items: 4 (avolition internal experience recreation), 10 (anhedonia past week frequency), 12 (lack of transitionary distress), 3 (Avolition Behavior Recreation), 8 (Asociality Internal Experience Social Media), 13 (Blunted Facial Affect), and 14 (Blunted Vocal Affect). IRT also suggested that reducing the number of anchor points from 0–6 to 0–5 would be ideal, as anchors 2 and 3 did not produce adequate separation.

Conclusions

Based on results of this study, we achieved the primary goals of evaluating the psychometric properties of the original 16-item version of the NSI-PR and creating a new final 11-item scale based on an iterative data-driven process. Table 11 highlights the key features and advances of this 11-item scale that was developed based on results of the current study. It will be important for future

Table 10. Scale Revision Summary of Changes from the 16-Item Beta Version to the Final 11-Item NSI-PR

(1)	Changed the anchor scaling from 0–6 to 0–5
(2)	Removal of the lack of transitionary distress item due to low factor communalities, high skew, poor inter-rater reliability, low item total correlation, IRT, and not being part of the 5 NIMH consensus domains
(3)	Removal of the social media internal experience item due to IRT and interviewer feedback. The probes/anchors for the asociality inner-experience item now include elements about social media into one inner-experience item that encompasses both in-person and electronic interactions
(4)	Removal of the anhedonia past week frequency item due to low factor loadings and conceptual overlap with avolition recreation behavior
(5)	Collapsing the 4 avolition items (in the beta version of the scale these were separated out into recreation and role for inner experience and overt behavior items) into 2 items that capture both role and recreation for inner-experience and behavior. This decision is based on based on IRT evidence about the avolition Recreation Behavior and Internal experience items functioning
(6)	Although blunted vocal and facial affect were identified for removal by IRT, they were not removed because they are considered core to the blunted affect construct. The anchors were revised in an effort to improve item function.

Note: Based on the combination of psychometric criteria reviewed in the results and discussion sections, several decisions were made to revise the 16-item beta NSI-PR and derive a final 11-item scale. The key decisions are reviewed above.

Table 11. Summary of Advances Offered by the Revised 11-Item Final NSI-PR

(1)	The final scale contains multiple materials designed to allow the NSI-PR to be used in multi-site CHR trials and clinical practice, including: a.a comprehensive manual; b.interview guide with suggested probes; c.score sheet and redcap forms; d.frequently asked questions document with rules and recommended procedures; e.professionally developed 1-h training video; f.multiple gold standard reliability videos with individuals displaying a range of symptom severity, gold standard ratings, and explanations for gold standard ratings that allow for extensive training and rater calibration within and across sites
(2)	The 11-item version and its associated materials are now being used in multiple studies/grants, by over 50 sites worldwide, including the Accelerating Medicines Partnership Program Schizophrenia (Amp-Scz) ProNet and PRESCIENT multi-site studies. The final 11-item scale has been translated into multiple languages (eg, Spanish, Korean, Italian, German, and Traditional Chinese script).
(3)	The final scale includes several highly innovative features that we believe will enhance its utility and psychometric value: a.Values used in anchors for avolition, anhedonia, and asociality items were created based on Ecological Momentary Assessment and social media data collected in conjunction with Study 1. The values tied to individual anchor ranges were based on quintile estimates using data related to motivation, social desire, anticipation of pleasure, experienced pleasure, frequency of social interaction, frequency of goal-directed activity, social media use, and text behavior (details to be reported in the subsequent validation study). These more objectively derived anchors were selected to be more psychometrically appropriate for the CHR population and capture the full range of normal to extreme levels of negative symptom pathology and avoid arbitrary designations made in older scales. b.Electronic media items that capture the type of social behavior that is highly common among today's youth; such behaviors are missed by older scales c.Avoids outdated conceptualizations and construct conflation that affect scales widely in use by the field d.Evaluates the 5 consensus domains that are emerging as the most valid conceptualization of the latent structure of negative symptoms across measures and phases of illness e.Isolates primary negative symptoms and avoids conflation of secondary negative symptom sources f.Isolates internal experience and behavior separately for avolition and asociality; the ability to observe dissociations between internal experience and behavior may be critical for observing treatment effects that often lead to behavioral change prior to internal experience.

investigations to study the validity of the 11-item finalized scale, including in more diverse samples (eg, more balanced sex distribution, greater racial diversity, and more participants recruited from acute settings), which is the goal of the second phase of this scale development process. By providing an improved outcome measure for assessing negative symptoms in CHR individuals, the NSI-PR may facilitate research into the mechanisms of negative symptoms and the development of novel therapeutic approaches. When used in conjunction with other clinical rating scales, behavioral measures, and computational approaches,[47] the NSI-PR may also facilitate the creation of novel risk calculator algorithms that have proved useful in longitudinal risk monitoring and diagnosis.[48–50]

Supplementary Material

Supplementary material is available at https://academic.oup.com/schizophreniabulletin/.

Funding

Research was funded by R01-MH116039 to Drs Strauss and Mittal, R01-MH 112545 to Dr Mittal, R01-MH-112613 to Dr Ellman, and R01-MH112612 to Dr Schiffman.

Conflict of Interest

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Foussias G, Mann S, Zakzanis KK, van Reekum R, Remington G. Motivational deficits as the central link to functioning in schizophrenia: a pilot study. *Schizophr Res.* 2009;115(2-3):333–337. doi:10.1016/j.schres.2009.09.020.
2. Luther L, Suor JH, Rosen C, Jobe TH, Faull RN, Harrow M. Clarifying the direction of impact of negative symptoms and neurocognition on prospective work functioning in psychosis: a 20-year longitudinal study. *Schizophr Res.* 2020;220:232–239. doi:10.1016/j.schres.2020.03.012.
3. Blanchard JJ, Kring AM, Horan WP, Gur R. Toward the next generation of negative symptom assessments: the collaboration to advance negative symptom assessment in schizophrenia. *Schizophr Bull.* 2011;37(2):291–299. doi:10.1093/schbul/sbq104.
4. Fusar-Poli P, Papanastasiou E, Stahl D, *et al*. Treatments of negative symptoms in schizophrenia: meta-analysis of 168 randomized placebo-controlled trials. *Schizophr Bull.* 2015;41(4):892–899. doi:10.1093/schbul/sbu170.
5. Kirkpatrick B, Fenton WS, Carpenter WT, Marder SR. The NIMH-MATRICS consensus statement on negative symptoms. *Schizophr Bull.* 2006;32(2):214–219. doi:10.1093/schbul/sbj053.
6. Kirkpatrick B, Strauss GP, Nguyen L, *et al*. The brief negative symptom scale: psychometric properties. *Schizophr Bull.* 2011;37(2):300–305. doi:10.1093/schbul/sbq059.
7. Kring AM, Gur RE, Blanchard JJ, Horan WP, Reise SP. The clinical assessment interview for negative symptoms (CAINS): final development and validation. *Am J Psychiatry.* 2013;170(2):165–172. doi:10.1176/appi.ajp.2012.12010109.
8. Galderisi S, Mucci A, Dollfus S, *et al*. EPA guidance on assessment of negative symptoms in schizophrenia. *Eur Psychiatry.* 2021;64(1):e23. doi:10.1192/j.eurpsy.2021.11.
9. Piskulic D, Addington J, Cadenhead KS, *et al*. Negative symptoms in individuals at clinical high risk of psychosis. *Psychiatry Res.* 2012;196(2-3):220–224. doi:10.1016/j.psychres.2012.02.018.
10. Zhang T, Tang X, Li H, *et al*. Clinical subtypes that predict conversion to psychosis: a canonical correlation analysis study from the ShangHai At Risk for Psychosis program. *Aust N Z J Psychiatry.* 2020;54(5):482–495. doi:10.1177/0004867419872248.
11. Carrión RE, Demmin D, Auther AM, *et al*. Duration of attenuated positive and negative symptoms in individuals at clinical high risk: associations with risk of conversion to psychosis and functional outcome. *J Psychiatr Res.* 2016;81:95–101. doi:10.1016/j.jpsychires.2016.06.021.
12. Yung AR, McGorry PD. The initial prodrome in psychosis: descriptive and qualitative aspects. *Aust N Z J Psychiatry.* 1996;30(5):587–599. doi:10.3109/00048679609062654.
13. Miller TJ, McGlashan TH, Rosen JL, *et al*. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull.* 2003;29(4):703–715. doi:10.1093/oxfordjournals.schbul.a007040.
14. Yung AR, Yung AR, Yuen HP, *et al*. Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states. *Aust N Z J Psychiatry.* 2005;39(11-12):964–971. doi:10.1080/j.1440-1614.2005.01714.x.
15. Strauss GP, Pelletier-Baldelli A, Visser KF, Walker EF, Mittal VA. A review of negative symptom assessment strategies in youth at clinical high-risk for psychosis. *Schizophr Res.* 2020;222:104–112. doi:10.1016/j.schres.2020.04.019.
16. Gur RE, March M, Calkins ME, *et al*. Negative symptoms in youths with psychosis spectrum features: complementary scales in relation to neurocognitive performance and function. *Schizophr Res.* 2015;166(1-3):322–327. doi:10.1016/j.schres.2015.05.037.
17. Strauss GP, Chapman HC. Preliminary psychometric properties of the brief Negative Symptom Scale in youth at Clinical High-Risk for psychosis. *Schizophr Res.* 2018;193:435–437. doi:10.1016/j.schres.2017.07.051.
18. First MB, Williams JB, Karg RS, Spitzer RL. *Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)*. Arlingt VA: Am Psychiatr Assoc; 2015:1–94.
19. Auther AM, Smith CW, Cornblatt B. *Global Functioning: Social Scale (GF: Social)*. Glen Oaks NY: Zucker-Hillside Hosp.; 2006.
20. Niendam TA, Bearden CE, Johnson JK, *et al*. Neurocognitive performance and functional disability in the psychosis prodrome. *Schizophr Res.* 2006;84(1):100–111. doi:10.1016/j.schres.2006.02.005.
21. Pelletier-Baldelli A, Strauss GP, Visser KH, Mittal VA. Initial development and preliminary psychometric properties of the Prodromal Inventory of Negative Symptoms (PINS). *Schizophr Res.* 2017;189:43–49. doi:10.1016/j.schres.2017.01.055.
22. Strauss GP, Hong LE, Gold JM, *et al*. Factor structure of the brief negative symptom scale. *Schizophr Res.* 2012;142(1-3):96–98. doi:10.1016/j.schres.2012.09.007.
23. Kring AM, Elis O. Emotion deficits in people with schizophrenia. *Annu Rev Clin Psychol.* 2013;9(1):409–433. doi:10.1146/annurev-clinpsy-050212-185538.
24. Strauss GP, Gold JM. A new perspective on anhedonia in schizophrenia. *Am J Psychiatry.* 2012;169(4):364–373. doi:10.1176/appi.ajp.2011.11030447.
25. Gruber J, Van Meter A, Gilbert KE, *et al*. Positive emotion specificity and mood symptoms in an adolescent outpatient sample. *Cogn Ther Res.* 2017;41(3):393–405. doi:10.1007/s10608-016-9796-7.
26. Jhung K, Park JY, Song YY, Kang JI, Lee E, An SK. Experiential pleasure deficits in the prodrome: a study of emotional experiences in individuals at ultra-high risk for psychosis and recent-onset schizophrenia. *Compr Psychiatry.* 2016;68:209–216. doi:10.1016/j.comppsych.2016.04.021.
27. Schlosser DA, Fisher M, Gard D, Fulford D, Loewy RL, Vinogradov S. Motivational deficits in individuals at-risk for psychosis and across the course of schizophrenia. *Schizophr Res.* 2014;158(1-3):52–57. doi:10.1016/j.schres.2014.06.024.
28. Strauss GP, Ruiz I, Visser KH, Crespo LP, Dickinson EK. Diminished Hedonic response in neuroleptic-free youth at

ultra high-risk for psychosis. *Schizophr Res Cogn*. 2018;12:1–7. doi:10.1016/j.scog.2017.12.001.

29. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychom Monogr Suppl.* 1969.
30. Chang WC, Strauss GP, Ahmed AO, *et al*. The latent structure of negative symptoms in individuals with attenuated psychosis syndrome and early psychosis: support for the 5 consensus domains. *Schizophr Bull.* 2021;47(2):386–394. doi:10.1093/schbul/sbaa129.
31. Rosseel Y. An R package for structural equation modeling. *J Stat Softw*. 2012;48(2). doi:10.18637/jss.v048.i02.
32. Asparouhov T, Muthén B. Structural equation models and mixture models with continuous nonnormal skewed distributions. *Struct Equ Model Multidiscip J.* 2016;23(1):1–19. doi:10.1080/10705511.2014.947375.
33. Hu LT, Bentler PM. Evaluating model fit. In: Hoyle RH, ed. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications, Inc; 1995:76–99.
34. Haguiara B, Koga G, Diniz E, *et al*. What is the best latent structure of negative symptoms in schizophrenia? A systematic review. *Schizophr Bull Open*. 2021;2(1):sgab013. doi:10.1093/schizbullopen/sgab013.
35. Strauss GP, Nuñez A, Ahmed AO, *et al*. The latent structure of negative symptoms in schizophrenia. *JAMA Psychiatry*. 2018;75(12):1271-1279. doi:10.1001/jamapsychiatry.2018.2475.
36. Dollfus S, Mucci A, Giordano GM, *et al*. European Validation of the Self-Evaluation of Negative Symptoms (SNS): a large multinational and multicenter study. *Front Psychiatry.* 2022;13:826465. doi:10.3389/fpsyt.2022.826465.
37. Havers L, Cardno A, Freeman D, Ronald A. The latent structure of negative symptoms in the general population in adolescence and emerging adulthood. *Schizophr Bull Open*. 2022;3(1):sgac009. doi:10.1093/schizbullopen/sgac009.
38. Ahmed AO, Kirkpatrick B, Galderisi S, *et al*. Cross-cultural validation of the 5-factor structure of negative symptoms in schizophrenia. *Schizophr Bull.* 2019;45(2):305–314. doi:10.1093/schbul/sby050.
39. Ang MS, Rekhi G, Lee J. Validation of the brief negative symptom scale and its association with functioning. *Schizophr Res.* 2019;208:97–104. doi:10.1016/j.schres.2019.04.005.
40. Jeakal E, Park K, Lee E, Strauss GP, Choi KH. Validation of the Brief Negative Symptom Scale in Korean patients with schizophrenia. *Asia-Pac Psychiatry*. 2020;12(3):e12382. doi:10.1111/appy.12382.
41. Mucci A, Galderisi S, Merlotti E, *et al*; Italian Network for Research on Psychoses. The Brief Negative Symptom Scale (BNSS): independent validation in a large sample of Italian patients with schizophrenia. *Eur Psychiatry.* 2015;30(5):641–647. doi:10.1016/j.eurpsy.2015.01.014.
42. Sun J, Ching EWK, Strauss GP, Ahmed AO, Chui WWH, Chow JSY. Validation of the traditional script Chinese version of the brief negative symptom scale. *Asian J Psychiatry.* 2021;55:102522. doi:10.1016/j.ajp.2020.102522.
43. Tam MHW, Ling-Ling W, Cheng K, *et al*. Latent structure of self-report negative symptoms in patients with schizophrenia: a preliminary study. *Asian J Psychiatry.* 2021;61:102680. doi:10.1016/j.ajp.2021.102680.
44. Turner JA, Calhoun VD, Thompson PM, *et al*. ENIGMA + COINSTAC: improving findability, accessibility, interoperability, and re-usability. *Neuroinformatics*. 2021;21:261-275. doi:10.1007/s12021-021-09559-y.
45. Ahmed AO, Kirkpatrick B, Granholm E, *et al*. Two factors, five factors, or both? External validation studies of negative symptom dimensions in schizophrenia. *Schizophr Bull.* 2022;48(3):620–630. doi:10.1093/schbul/sbab148.
46. Strauss GP, Esfahlani FZ, Galderisi S, *et al*. Network analysis reveals the latent structure of negative symptoms in schizophrenia. *Schizophr Bull.* 2019;45(5):1033–1041. doi:10.1093/schbul/sby133.
47. Gold JM, Corlett PR, Strauss GP, *et al*. Enhancing psychosis risk prediction through computational cognitive neuroscience. *Schizophr Bull.* 2020;46(6):1346–1352. doi:10.1093/schbul/sbaa091.
48. Cannon TD, Yu C, Addington J, *et al*. An individualized risk calculator for research in prodromal psychosis. *Am J Psychiatry.* 2016;173(10):980–988. doi:10.1176/appi.ajp.2016.15070890.
49. Fusar-Poli P, Rutigliano G, Stahl D, *et al*. Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*. 2017;74(5):493–500. doi:10.1001/jamapsychiatry.2017.0284.
50. Zhang T, Xu L, Tang Y, *et al*; SHARP (ShangHai At Risk for Psychosis) Study Group. Prediction of psychosis in prodrome: development and validation of a simple, personalized risk calculator. *Psychol Med.* 2019;49(12):1990–1998. doi:10.1017/S0033291718002738.