Development of a Scoring Guide for the Positive and Negative Syndrome Scale (PANSS) Abstract Thinking Subscale

Luba Leontieva[1*],[Jackie Dimmock][1], Kate Carey[2], Robert J Ploutz-Snyder[3], Zsuzsa Szombathyne Meszaros[1] and Steven L Batki[4]

[1]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA

[2]Brown University, Providence RI, USA

[3]University of Michigan School of Nursing, Ann Arbor, MI, USA

[4]Department of Psychiatry, UCSF; and San Francisco VA Medical Center, San Francisco, CA, USA

[*]**Corresponding author**: Luba Leontieva, Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA, Tel: + (315) 464-3278; E-mail: leontiel@upstate.edu

Abstract

Background: The Positive and Negative Syndrome Scale (PANSS) is a widely used instrument for symptom severity assessment in schizophrenia. Its Abstract Thinking item (N5) was developed for the assessment of thought disorder. This item currently lacks examples of correct and incorrect responses to similarities and proverb items. Different raters may judge these items as correct or incorrect based on their own level of abstraction, cultural background, and familiarity with possible responses. Precision in scoring is especially important, when the instrument is used to evaluate changes in schizophrenia symptoms over time and with treatment. This study proposes a new method of scoring the N5 subscale.

Objectives: The aim of the present study was to develop a new scoring guide for the PANSS N5 Similarities and Proverbs scale and to assess inter-rater reliability using the newly developed scoring guide.

Method: The authors analysed responses to PANSS questions of subjects who completed a double blind, randomized, placebo controlled clinical trial of oral naltrexone for treatment of alcohol use disorders in schizophrenia. Results: Of the 90 subjects, 45 had schizophrenia and 45 had schizoaffective disorder. 95% of subjects had alcohol dependence, 5% had alcohol abuse. Subjects consumed a median of 21 standard drinks per week at study entry. Participants had low to moderate PANSS Positive, Negative, and General Psychopathology scores. 434 different responses to similarities and 748 different responses to proverbs were sorted independently by two psychologists using a newly developed scoring guide for N5. The guide sorted responses into 4 categories, from correct to marginal, to concrete, to incorrect; examples of almost each type of responses were provided in the guide. Inter-rater

reliability for scoring all Similarities responses was 93%, Weighed Cohen Kappa 0.83, p<0.001; for scoring all Proverbs was 87%, Weighted Cohen Kappa 0.62, p<0.001.

Conclusion: Strong inter-rater reliability was achieved using a newly-developed scoring guide for Similarities and Proverbs of PANSS. The Guide could be used to improve accuracy of scoring PANSS N5.

Keywords: Schizophrenia; Abstract thinking; Similarities; Proverb interpretations; PANSS

Introduction

Difficulty with abstract thinking is considered a core symptom of schizophrenia since Bleuler. Goldstein noted that the thought process of individuals with schizophrenia included maladaptive, concrete thinking. According to Benjamin, patients with schizophrenia could not infer abstract meaning from proverbs, and used literal interpretation instead. Gorham showed significant drop between healthy controls and patients with schizophrenia in abstract scores and rise in concrete scores, indicating severe impairment in abstract thinking in patients with schizophrenia [1-4].

An understanding of figurative language and an ability to use abstraction is crucial in successful interpersonal interaction, which patients with schizophrenia are lacking. Positive changes in abstract thinking during treatment may indicate improvement in interpersonal functioning in these individuals, and it is important to have an instrument to precisely assess those changes.

The Positive and Negative Syndrome Scale (PANSS) is a widely used scale to assess symptoms of schizophrenia. The example of it wide use could be the fact that Kay's original PANSS article was sited in 3,020 articles in United States National Library of Medicine. It has established psychometric

properties and is broadly used in research. The history of the development of PANSS originated in the need to have a valid and reliable measure of positive and negative symptoms and its changes with treatment. At the time of PANSS development there were lack of such scales. PANSS consists of Positive Symptoms scale which evaluates delusions, hallucinations, and disordered thinking and Negative Symptoms Scale which evaluated such negative symptoms as apathy and avolition [5-7].

The Abstract Thinking subscale (N5) of PANSS consists of 32 items: 16 similarities and 16 proverbs. Both similarities and proverbs are divided in blocks of four, with easier items at the beginning and more difficult items towards the end. On a single administration, a patient is usually asked four similarities and four proverbs, one from each block. On repeated administration, the items are systematically rotated to minimize repetition. Scoring is based on responses to similarities and proverb interpretation and on the assessment of concrete versus abstract thinking by the interviewer.

Based on the PANSS manual, scoring of the Abstract Thinking subscale involves a 7-point scale, ranging from 1 (no difficulties or perfect score) to 7 (no items are answered correctly on similarities/proverbs and patient cannot interact even minimally due to severe impairment). Scores in the range of 2 to 6 represent different degrees of abstract thinking impairment, ranging from minimal/questionable pathology (score of 2) to severe (score of 6). The rating is assigned based on an outline in the scoring manual; for example, a rating of 3 is assigned when the patient gives a literal or personalized interpretation to more difficult proverbs and categories. No examples are given in the manual of what constitutes a "literal/personalized interpretation" or other key coding constructs.

Depending on the examiner's ability for abstraction, cultural background, and familiarity with the proverbs, the patient's responses could be judged differently. Proverbs reflect the succinct wisdom of the culture from which the particular proverb is originated [8]. A more systematic way of assessing and scoring responses to similarities and proverbs could be helpful to minimize the error in interpretation, ease the scoring, and provide a valid tool in assessment the Abstract Thinking Item of the PANSS. This is particularly important if the instrument is used repeatedly to evaluate effect of therapy on abstract thinking. Changes in this area, which could be an important indication of effectiveness of treatment, might be missed due to imprecise assessment.

The aim of the present research study was to develop a new scoring guide that can be used in scoring of Abstract Thinking (N5) subscale items of PANSS and evaluate its inter-rater reliability. The authors will outline the methodology of developing the scoring guide and report inter-rater reliability data using this guide.

Methodology

This report analyses data obtained interviewing 90 participants who completed a NIAAA-funded, double-blind, randomized, placebo controlled clinical trial of directly observed oral naltrexone for treatment of alcohol use disorders in schizophrenia. Data were collected from November 2003 to June 2008. All patients were native English speakers, carried the diagnosis of schizophrenia or schizoaffective disorder. The patients were recruited from several outpatient mental health clinics, from central New York, the city of the population of approximately 147,000 people. Demographic and clinical characteristics of study participants are summarized in **Table 1**. More detailed clinical characteristics of study patients reflected elsewhere [9,10].

Table 1: Demographic and clinical characteristics of study participants.

Demographics	N	%
Gender: Male	64	71
Race: African American Caucasian	37 38	41 42
Marital status: never married Divorced, separated	58 25	64 28
Unemployed	81	90
Other demographics	Mean (SD)	Range
Age (Years)	42 (9)	22-59
Education (years)	12 (2)	6-17
Monthly income	$ 417 ($ 351)	$0 - 1300
Clinical characteristcics	Mean (SD)	Range
PANSS General	32.4 + 7.1	20-49
PANSS Positive Symptoms	15.4 + 5.2	7-31
PANSS Negative Symptoms	13.5 + 4.9	7-30
Diagnostic characteristics	N	%
Schizophrenia	45	50
Schizoaffective disorder	45	50
Alcohol dependence	86	95
Alcohol abuse	4	5

The scoring guide was developed based on the PANSS N5 rating scale. Examples of concrete and abstract answers were collected from patients with schizophrenia/schizoaffective disorder and alcohol use disorders, who participated in another study as well as from patients with schizophrenia during routine clinical interviews with patients at state psychiatric hospital. The examples were rated according to PANSS N5 rating scale by two PANNS Institute-certified psychologists. The cultural backgrounds of the psychologists were as follows: JD was native of the USA, native English speaker; LL was Russian-born, fluent English speaker, American-educated and fully acculturated. Both psychologists discussed and assigned the ratings of these examples together. The correct definitions of the proverbs were taken from the proverbs dictionary. The dictionary is the most comprehensive

North American proverbs and saying to date and is differ from other such dictionaries in that they were used and originated in America; they are relatively modern - collected from thousands of books, journals, magazines, and newspapers published between 1880's-1990s [7,11,12].

The responses to Abstract Thinking items were first transcribed from the PANSS administration form. Altogether 1182 different responses (excluding responses such as "I do not know") were printed on two sets of index cards. Two PhD-level psychologists, who are PANSS Institute-certified raters, mentioned above, independently sorted the responses into separate and mutually exclusive categories using the scoring guide (Appendix 1 and 2).

All statistical analyses were conducted using Stata, IC. Inter-rater reliability was evaluated using weighted Cohen's Kappa statistic such that perfect agreement between the two raters was assigned a weight of 1; disagreements by only 1 unit were weighted as 0.75; disagreements by 2 units received a weight of 0.50; and disagreements by more than 2 units received zero weight. These weights were chosen based on our clinical assessment that perfect agreement is best, with linearly less value assigned to items that resulted in two expert raters assigning scores that differed by 1 or 2 units, and our conclusion that if raters' scores disagree by 3 or more, then we ascribe zero worth to that level of discordance. Agreement is reported according to this weight scheme, with 2-tailed alpha to reject the null hypothesis set at 0.05.

Results

Overall agreement on Similarities was 93%, Weighted Cohen's Kappa=0.826, p<0.0001. Graphical representation of percents agreement between two raters on Similarities items can be seen in **Figure 1**. The poorest agreement rate was on Table & Chair (64% agreement).
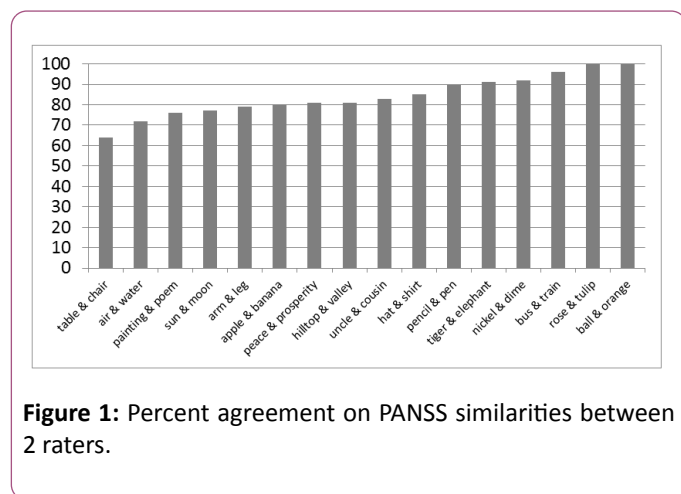

Figure 1: Percent agreement on PANSS similarities between 2 raters.

The perfect agreement rate was on Ball & Orange and Rose & Tulip (100% respectively). Among 16 similarities, even relatively abstract pairs such as "Peace and prosperity" and "The sun and the moon" attained 80% correspondence between two raters. In Similarities, the majority of differences were between rating (4-5) and (6-7). For example, the

following responses on this similarity pair were scored differently for Table & Chair: "Both supports," "Both used to eat," "Both stands," "Both for eating."

Overall agreement for 16 Proverbs was 87%, Weighted Cohen's Kappa=0.616, p<0.0001. Graphical representation of per cents agreement between two raters on Proverbs items can be seen in **Figure 2**. The poorest agreement was on "Too many cooks spoil the broth" (26%); the best agreement was on "A rolling stone gathers no moss" (88%).
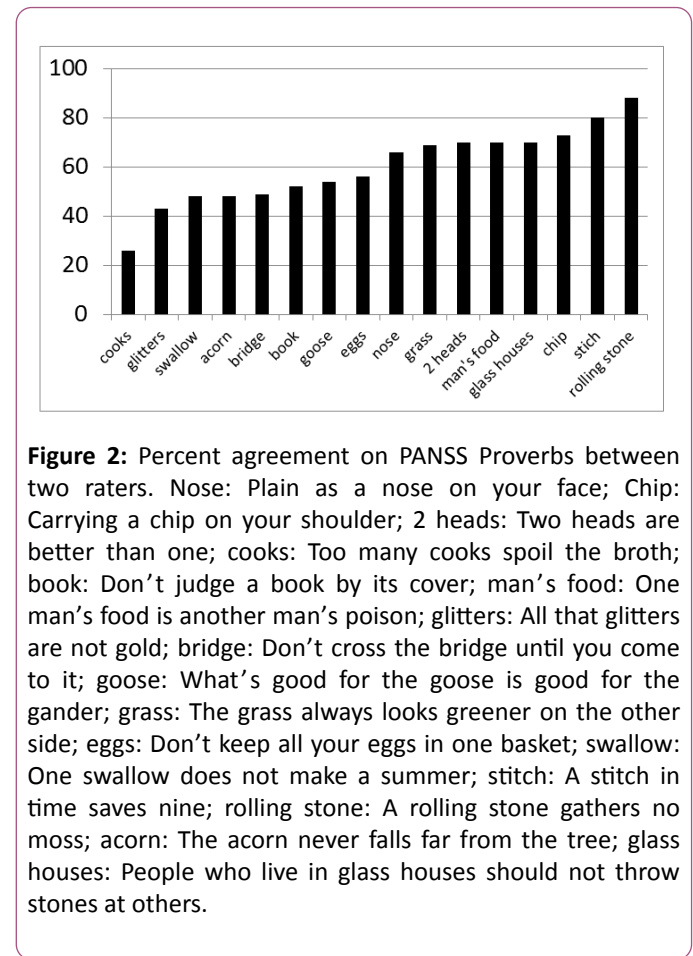

Figure 2: Percent agreement on PANSS Proverbs between two raters. Nose: Plain as a nose on your face; Chip: Carrying a chip on your shoulder; 2 heads: Two heads are better than one; cooks: Too many cooks spoil the broth; book: Don't judge a book by its cover; man's food: One man's food is another man's poison; glitters: All that glitters are not gold; bridge: Don't cross the bridge until you come to it; goose: What's good for the goose is good for the gander; grass: The grass always looks greener on the other side; eggs: Don't keep all your eggs in one basket; swallow: One swallow does not make a summer; stitch: A stitch in time saves nine; rolling stone: A rolling stone gathers no moss; acorn: The acorn never falls far from the tree; glass houses: People who live in glass houses should not throw stones at others.

Discussion

The accuracy of PANSS scoring has been discussed in comprehensive systematic review studies. Because PANSS is so widely used instrument in schizophrenia research, especially in medical trials, it is important to have a reliable and consistent across scorers guide to score PANSS N5 correctly. The more accurate differences in abstract thinking could point out on the success or failure of newly developed medications that could improve cognitive functioning. This paper reports an inter-rater reliability for a newly developed scoring guide of the PANSS N5 subscale: "Difficulty in abstract thinking." This is the first scoring guide developed to improve accuracy of scoring similarities and proverbs of the PANSS [13,14].

The scoring guide achieved strong inter-rater reliability of both similarities and proverbs.

Our findings point out that in Similarities, more precise definition of the concrete rating (rating 4-5) is needed for some similarity pairs, for example: functional aspect of each word is used but no main similarity between words is reported.

The scoring guide reached an excellent inter-rater reliability for proverbs. This is especially remarkable as proverbs are difficult to judge due to multiple possible meanings. The highest agreement (88%) was achieved for the proverb "A rolling stone gathers no moss" which is one of the most difficult proverbs in the set. As in the case of similarities, most of the disagreements were between 4-5 and 6-7 ratings. Some examples of the answers were: "Too many opinions," "Too many trying to be in charge," "Too many hands doing the same things." To improve the rating of this proverb more attention should be paid to the complete meaning of the response phrase. For example, response "Too many opinions" represents an incomplete response with some degree of abstraction and should be scored as 4-5 if no further elaboration is offered by the responder.

The newly developed scoring guide may be used to achieve more accurate scoring of the N5 item compared to routine administration of the PANSS scale. Scoring of similarities and proverbs could be the primary scoring of N5 in the future, while "concrete vs. abstract mode of the interview" could be a secondary or complementary score as it includes subjective view of the examiner based on the interview. The precise changes in abstract thinking in patients with treatment can point out on the improvement of figurative language comprehension, which is crucial for successful social interaction [15]. Also, specific cognitive remediation programs could be developed to improve figurative language interpretation to help patients deal with ambiguous social stimuli [16]. The efficacy of such programs could be measured with more precise PANSS N5 scoring using the new guide.

Conclusion

One of the study limitations is that we could not find examples for some of the responses. Most of these absent examples are ratings of 2-3 on similarities. This could be explained by the fact that our sample consisted of individuals with schizophrenia who had difficulties with abstractions. So, they either got the target item right or responded in concrete or incorrect fashion. Another limitation is that on the similarities "The Sun and the Moon" we included the answer "Planets" as a correct response. More accurate answer would be "celestial bodies." However, we obtained responses from several people without mental illness, asking to provide the similarity between the sun and the moon and a majority of those asked stated that they were planets. Thus, we felt that such response represents the most common interpretation of the similarity between these items.

This new guide will be further tested, published, and made available for researchers. It is strongly recommended to use it to improve PANSS scoring.

References

1. Bleuler E (1950) Dementia Praecox or the Group of Schizophrenias. New York: International Universities Press.

2. Goldstein K (1959) Concerning the concreteness in schizophrenia. J Abnormal Social Psychol 59: 146-148.

3. Benjamin JD (1944) A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In: Kasanin JS (ed.) Language and Thought in Schizophrenia, Berkley, Calif.: University of California Press, pp: 65-90.

4. Gorham DR (1956) Use of the Proverbs Test for Differentiating Schizophrenics from Normals. J Consulting Psychol, 20: 435-440.

5. Kay SR, Fiszbein A Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 13: 261-276.

6. Liechti S, Capodilupo G, Opler DJ, Opler M, Yang LH (2017) A Developmental History of the Positive and Negative Syndrome Scale (PANSS). Innovations in Clinical Neuroscience 14: 12-17.

7. Kay SR, Opler LA, Fiszbein A (2006) Positive and Negative Syndrome Scale (PANSS) Technical Manual. MHS, North Tonawanda, New York.

8. HOU Rong (2013) Proverbs Reveal Culture Diversity. Cross Cultural Communication 9:31-35.

9. Batki SL, Dimmock JA, Ploutz-Snyder R, Meszaros ZS, Canfield K (2009) Directly Monitored Naltrexone Reduces Heavy Drinking in Schizophrenia: Preliminary Analysis of a Controlled Trial. Alcohol Clin Exp Res 33:1-346.

10. Meszaros ZS, Dimmock JA, Ploutz-Snyder RJ, Abdul-Malak Y, Leontieva L, et al. (2011) Predictors of smoking severity in patients with schizophrenia and alcohol use disorders. Am J Addict 20: 462-467.

11. Batki SL, Dimmock JA, Meszaros ZS, Ploutz-Snyder R, Cavallerano M, et al. (2010) Extended-Release Naltrexone for Alcohol Dependence in Serious Mental Illness. Alcohol Clin Exp Res 34: 661-176A.

12. Titelman GY (1996) Random House Dictionary of Popular Proverbs and Sayings. First Edition. Random House.

13. StataCorp (2011) Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

14. Obermeier M, Schennach-Wolf R, Meyer S, Möller H, Riedel M, etal. (2011) BMC Psychiatry 11:113.

15. Vallance DD, Wintre MG (1997) Discourse process underlying social competence in children with language learning disabilities. Development and Psychology 9: 95-108

16. Thoma P, Daum I (2006) Neurocognitive mechanisms of figurative language processing - evidence from clinical dysfunctions. Neurosci Behav Rev 30: 1182-1205.