Methodological considerations regarding the association of Stroop and verbal fluency performance with the symptoms of schizophrenia

Todd S. Woodwarda,b,*, Christian C. Ruffa,b,c, Allen E. Thorntonb,d, Steffen Moritza,b,e, Peter F. Liddlef,g

aDepartment of Psychology, University of British Columbia, Vancouver, BC, Canada
bDepartment of Medicine and Research, Riverview Hospital, Port Coquitlam, BC, Canada
cDepartment of Psychology, University of Freiburg, Freiburg, Germany
dDepartment of Psychology, Simon Fraser University, Burnaby, BC, Canada
eUniversity Hospital Hamburg-Eppendorf, Hamburg, Germany
fDepartment of Psychiatry, University of British Columbia, Vancouver, BC, Canada
gDepartment of Psychiatry, University of Nottingham, Nottingham, UK

Received 11 October 2001; received in revised form 11 January 2002; accepted 16 January 2002

Abstract

Previous research on schizophrenia has reported conflicting findings regarding the association between Stroop performance and the disorganization syndrome, as well as performance on verbal fluency tests and the psychomotor poverty syndrome. In the present work, we consider whether these inconsistencies may be increased due to variations in test format and failures to report the appropriate test parameters. In 36 schizophrenic inpatients, we administered list and single-trial versions of the Stroop test, and report the correlation with the disorganization syndrome for both errors and speed. For verbal fluency, we separated the total score into measures of switching and clustering, and observed the relationship with psychomotor poverty. For both versions of the Stroop test, accuracy, but not speed, was correlated with disorganization. For verbal fluency, decreased cluster production relative to total words generated was associated with psychomotor poverty, but the number of switches between clusters was not. It is suggested that assessing and reporting a full range of test parameters can reduce between-study inconsistencies. Cognitive interpretations for the present set of results are discussed.

Keywords: Schizophrenia; Stroop test; Verbal fluency test; Psychomotor poverty syndrome; Disorganization syndrome

1. Introduction

Early theoretical models have suggested that the symptoms of schizophrenia can be split into positive and negative types (Andreasen, 1984a,b; Crow, 1980). However, more recent evidence suggests that schizophrenia is composed of three distinct syndromes...
Arndt et al., 1991; Bilder et al., 1985; Liddle, 1987). Liddle (1987) labelled these three syndromes as reality distortion (hallucinations and delusions), disorganization (thought disorder and inappropriate affect) and psychomotor poverty (flat affect, reduction in spontaneous movement and poverty of speech). The three-syndrome account also differs from the positive-negative dichotomy in that it is a dimensional model: one patient can display symptoms from a combination of the three syndromes.

Several cognitive tasks have been linked to the symptom dimensions of schizophrenia (e.g., O’Leary et al., 2000), providing insight into the nature of the underlying brain systems (George and Neufeld, 1985; Mortimer, 1992). Psychomotor poverty and disorganization, but not reality distortion, are commonly observed to correlate negatively with performance on cognitive tasks. The disorganization syndrome is most often associated with problems in shifting set, and in overcoming Stroop interference (e.g., Baxter and Liddle, 1998; Liddle and Morris, 1991; Moritz et al., 2001a). Psychomotor poverty, on the other hand, tends to be associated with verbal fluency and memory measures (Aleman et al., 1999; Liddle and Morris, 1991; Moritz et al., 2001a). Nonetheless, despite a body of literature which has produced these associations between cognitive tasks and the symptom dimensions of schizophrenia, several studies have produced equivocal results (see below).

In the present work, we will consider factors which may contribute to the apparent inconsistencies across studies with respect to the association between symptom dimensions, the Stroop test and verbal fluency tests. The Stroop test involves naming the ink color of letters forming a color word, when the meaning of the word is incongruent with the ink color. Administration can be either single-trial or list format, and either speed or errors (or both) can be measured. For the list version of the Stroop test, the association with disorganization has been reported only for speed, and both significant (Baxter and Liddle, 1998, experiment 1; Liddle and Morris, 1991; Moritz et al., 2001a) and nonsignificant (Hoschel et al., 1998; Ngan and Liddle, 2000; O’Leary et al., 2000) correlations have been reported. For the single-trial version, correlations between disorganization and both speed and errors can be found in the literature. In one study, both were found to be significantly correlated with disorganization (Barch et al., 1999b). In another study, only the correlation with errors was reported, and this was again significant (Barch et al., 1999a). In contrast, nonsignificant correlations have also been reported for speed (Baxter and Liddle, 1998, experiments 2 and 3). To directly compare the association of both speed and accuracy derived from both versions of the task, we administered both the single-trial and list versions of the Stroop test to one sample, and report correlations between disorganization and both speed and errors. To ensure that selective attention, and not color naming speed, underlies an association between Stroop speed and disorganization, we presented a block of neutral color naming trials for the single-trial version; exclusion of the confound is achieved by demonstrating that an association between incongruent-color-naming speed and disorganization remains significant when variance due to neutral-color-naming speed is subtracted out.

A common test of verbal fluency involves producing as many words as possible which meet a specific criterion (e.g., words beginning with the letter A). Although verbal fluency is often noted to correlate significantly with negative symptoms or psychomotor poverty (Frith et al., 1991; Hammer et al., 1995; Himelhoch et al., 1996, on neuroleptics; Liddle and Morris, 1991; Moritz et al., 2001a; Norman et al., 1997; O’Leary et al., 2000), some studies report no correlation (e.g., Himelhoch et al., 1996, off neuroleptics; Hornstein et al., 1998; Morrison-Stewart et al., 1992). One reason for this inconsistency is that verbal fluency performance can be affected by a number of cognitive variables, only some of which may correlate with psychomotor poverty. Two of these have been previously identified and operationalized: switching between mental sets, referred to as switching, and elaborating on selected mental sets, referred to as clustering. Switching and clustering contribute unique sources of variance to verbal fluency performance (Troyer et al., 1997). If only one of these variables is, in fact, associated with psychomotor poverty, reporting only the total number of words produced may obscure the critical association, and lead to the negative results reported above. Consequently, we have attempted to determine whether clustering, switching, or both, mediate the relationship between psychomotor poverty and verbal fluency.
In order to further elucidate the etiological significance of decreased verbal fluency in relation to psychomotor poverty, we have derived indices that control for overall word production. This procedure was adopted because the verbal fluency measure of total number of words produced may simply be another method of measuring poverty of speech, not a measure of a cognitive subcomponent which underlies psychomotor poverty. This reasoning extends to the measurement of switches and clusters. If total words produced is decreased due to the presence of a particular symptom (e.g., poverty of speech), it is redundant to report that the number of switches and clusters was also correlated with this symptom. This potential confound can be accounted for by dividing the count of switches and clusters by the total number of words produced; in this fashion, switching and clustering can be interpreted separately from speech production.

An additional problem common to all studies of the association between cognition and symptomatology in schizophrenia is that scales such as the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1989) and The Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1984a) measure disorganized and impoverished mental activity within single items. For example, SANS items such as “imper sistence at work” and “social inattentiveness” may be influenced by disorganized aspects of schizophrenia, such as disorganization of thought. Similarly, the PANSS item “blunted affect” includes “inappropriate uncontrolled laughter,” which would be better described as inappropriate affect, an aspect of the disorganization syndrome. As a final example, the PANSS item “disturbance of volition” is heavily influenced by decision-making ability, which can be affected by both disorganized and impoverished mental activity. If the negative and disorganized syndrome constructs share variance, correlations with cognitive tests could be attenuated due to splitting of the effect between the two syndromes. A short-interview rating scale assessing, and clearly separating these three syndromes has been developed, entitled The Signs and Symptoms of Psychotic Illness (SSPI; Liddle et al., 2003). This scale was utilized in the present study. In order to avoid the study of medication-induced slowing, only patients medicated with atypical neuroleptics were included.

2. Methods

2.1. Participants

Thirty-six patients (27 males, 9 females) diagnosed with Schizophrenia according to DSM-IV criteria were recruited from Riverview Hospital, Port Coquitlam, Canada. The mean age was 35.6 years (S.D. = 9.43) and the mean years of education was 12.8 (S.D. = 2.72). The mean duration of illness, since first hospitalization, was 10.22 years (S.D. = 10.13). The mean SSPI total score was 19.44, indicating moderate overall severity of psychotic illness. Patients were excluded if they met one of the following criteria: composite IQ (as measured by the K-BIT; Kaufman and Kaufman, 1990) of less than 70 (M = 95.8, S.D. = 11.36), history of acquired brain damage or traumatic head injury (e.g., with a loss of consciousness for more than 10 min), epilepsy, encephalitis, diabetes, hepatitis C or hypothyroidism. Patients with an Axis I diagnosis in addition to schizophrenia (e.g., polysubstance abuse) were also excluded. Finally, the primary language of all patients was English, their color vision was intact and their eyesight was 20/40 or better, both measured by the MIS Pocket Vision Guide. At the time of testing, all patients were receiving atypical neuroleptic medication. The experimental procedure was approved by the University of British Columbia and Riverview Hospital’s ethical review boards and all subjects gave written, informed consent.

2.2. Measures

Psychopathology was assessed using the Signs and Symptoms of Psychotic Illness rating scale (SSPI; Liddle et al., 2003). The SSPI is a 20-item, five-point rating scale, which can be completed after a 25–30-min semistructured interview with 15 direct questions about symptoms. The severity of each item is rated in the range of 0–4. The SSPI provides specific guidelines for the severity level of each item, but generally, a score of 1 denotes questionable abnormality, 2 denotes definite but mild abnormality, 3 denotes pathology of moderate severity that has a substantial impact on mental functioning and 4 indicates severe psychopathology. The following core items from the SSPI were used to quantify the
three-syndrome model of psychotic illness: Underactivity, Flattened Affect and Poverty of Speech for psychomotor poverty; Inappropriate Affect and Disordered Form of Thought for disorganization; and Delusions and Hallucinations for reality distortion. Symptom ratings were carried out blindly by TSW. Twenty-one of the interviews were videotaped and reviewed by both TSW and CCR. Syndrome aggregate score interrater reliability between CCR and TSW were as follows: Reality Distortion ($r = 0.83$), Psychomotor Poverty: ($r = 0.76$), Disorganization ($r = 0.75$).

2.2.1. List version of the Stroop test

For the list version of the Stroop Test (Trenerry et al., 1989), patients were given a list of 112 incongruent Stroop stimuli (Stroop-sheet CW), and were told: “I would like you to name aloud the colour of the ink—red, blue, green or tan—in which the word is printed. Go as quickly as you can, going down the columns. For this first one you would say ‘RED.’ Understand? If you make a mistake, just correct yourself and keep on going. Name the colour of the ink as quickly and as accurately as you can. Ready? Begin.” The subject was stopped after 120 s or the end of the list. Number of items completed was used as the measure of speed and errors were recorded on the response sheet.

2.2.2. Single-trial computerized Stroop test

For these experiments, separate blocks of neutral and incongruent Stroop stimuli (font uppercase Geneva 36 point) were presented on a 14-in. screen, using a PC controlled by Superlab software (Cedrus, 1999). Neutral Stroop stimuli were a display of XXXX displayed in blue, green, red or yellow. Incongruent Stroop stimuli were color names blue, green, red and yellow printed such that the color of the ink was incongruent with the color name. To create the two stimulus presentation files (neutral and incongruent), 25 incongruent Stroop stimuli were selected randomly (with replacement) from the 12 possible incongruent ink/word combinations, and 25 neutral Stroop stimuli were selected randomly (with replacement) from the four possible colors. Presentation order of the incongruent stimulus block (25 trials) and the neutral stimulus block (25 trials) was counterbalanced across subjects.

All stimuli were presented at the center of the screen. Prior to commencement of each block of trials, subjects were instructed to name the color and to respond as quickly and as accurately as possible. Vocal reaction time (RT) in milliseconds was recorded by Superlab in response to microphone activation. After the microphone was activated, a blank screen was presented, the experimenter typed a “c” or “e” to specify whether the correct response or an error was given, and then initiated advancement to the next trial with a key press. The measures of interest were trial-based mean RT in the incongruent block (excluding error trials), total number of errors in the incongruent block, and the Stroop interference score (RT in the incongruent block—RT in the neutral block).

2.2.3. Verbal fluency

Verbal fluency was assessed by the Controlled Oral Word Association Test (Spreen and Benton, 1969, 1977). Patients received the following instructions: “I will say a letter of the alphabet. Then I want you to give me as many words that begin with that letter as quickly as you can. For instance, if I say ‘B,’ you might give me ‘bad,’ ‘battle,’ ‘bed’... I do not want you to use words that are proper names such as ‘Boston,’ ‘Bob,’ or ‘Brylcreem.’ Also, do not use the same word again with a different ending such as ‘eat’ and ‘eating.’ Any questions? Begin when I say the letter. The first letter is ‘F.’ Go ahead.”

One minute was allowed for each letter (F, A and S). If patients discontinued before the end of the minute, they were encouraged to try to think of more words. If there was an initial silence of 15 s, the basic instructions and the letter were repeated. Clustering (first letters, rhymes, first and last sounds and homonyms) and switching between clusters were coded according to previously published criteria (Troyer et al., 1997). The measures of interest were total number of words produced for all letters (total score), as well as the total number of switches and clusters.

Tests were administered in two separate testing sessions, on separate days, as part of a larger study. In order of administration, the K-BIT, SSPI and Single-Trial Computerized Stroop Test were administered in session one, and the list version of the Stroop test and verbal fluency were administered in session two.
3. Results

The minimum criterion of three subjects to one variable (Pawlik, 1976) was met for the factor analysis reported below. Symptom ratings for the items Delusions, Hallucinations, Underactivity, Flattened Affect, Inappropriate Affect, Poverty of Speech and Disordered Form of Thought were submitted to a principal component analysis with a varimax rotation. In accordance with prior theory (e.g., Liddle, 1987), three factors were extracted. The results are listed in Table 1. As predicted, three factors corresponding to psychomotor poverty, disorganization and reality distortion emerged. The psychomotor poverty and disorganization factor scores were saved for the subsequent correlations with memory measures.

Because decreased performance was expected to be associated with increased symptoms, significance of Pearson correlations was assessed with one-tailed tests. The cut-off for significance was set to \( \alpha = 0.05 \). The number of items completed in the Stroop list version \((M = 84.37, \text{S.D.} = 20.39)\) was not significantly correlated with disorganization, \( r(33) = -0.03, p = 0.44 \). Similarly, neither the average RT for single-trial Stroop color naming \((M = 922.50, \text{S.D.} = 191.76)\), nor the Stoop interference index \((M = 195.25, \text{S.D.} = 123.48)\) was significantly correlated with disorganization; \( r(34) = 0.05, p = 0.38, r(34) = 0.09, p = 0.31 \), respectively. However, errors for both the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Rotated component matrix for Signs and Symptoms of Psychotic Illness items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Psychomotor poverty</td>
</tr>
<tr>
<td>Underactivity</td>
<td>0.83</td>
</tr>
<tr>
<td>Flattened affect</td>
<td>0.92</td>
</tr>
<tr>
<td>Poverty of speech</td>
<td>0.76</td>
</tr>
<tr>
<td>Inappropriate affect</td>
<td>-0.24</td>
</tr>
<tr>
<td>Disordered form of thought</td>
<td>-0.06</td>
</tr>
<tr>
<td>Delusions</td>
<td>-0.26</td>
</tr>
<tr>
<td>Hallucinations</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

Extraction method: Principal Component Analysis. Rotation method: Varimax with Kaiser Normalization. All loadings above 0.50 are set in bold font.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Pearson Correlations between cognitive tests and syndrome scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Psychomotor poverty</td>
</tr>
<tr>
<td>List Stroop errors*</td>
<td>-0.13</td>
</tr>
<tr>
<td>List Stroop speed*</td>
<td>-0.14</td>
</tr>
<tr>
<td>Single-trial Stroop errors</td>
<td>-0.12</td>
</tr>
<tr>
<td>Single-trial Stroop interference</td>
<td>-0.07</td>
</tr>
<tr>
<td>Single-trial Stroop speed</td>
<td>-0.09</td>
</tr>
<tr>
<td>Verbal fluency total words</td>
<td>0.00</td>
</tr>
<tr>
<td>Verbal fluency switches</td>
<td>0.02</td>
</tr>
<tr>
<td>Verbal fluency switches (total)</td>
<td>0.04</td>
</tr>
<tr>
<td>Verbal fluency clusters</td>
<td>-0.27</td>
</tr>
<tr>
<td>Verbal fluency clusters (total)</td>
<td>-0.42 **</td>
</tr>
</tbody>
</table>

* One subject did not complete the list version of the Stroop test (\(N = 35\)) and five subjects reached ceiling (112 items).

\( p = 0.06 \).

\( * p \leq 0.05 \).

\( ** p \leq 0.001 \) (one-tailed).

List \((M = 0.78, \text{S.D.} = 1.22)\) and the single-trial incongruent condition \((M = 1.28, \text{S.D.} = 1.47)\) were significantly correlated with disorganization, \( r(33) = 0.28, p = 0.05; r(34) = 0.34, p = 0.02 \), respectively. Reality distortion was not significantly correlated with any measure on either test version (see Table 2).

In terms of the verbal fluency results, neither total score \((M = 36.61, \text{S.D.} = 11.96)\) nor number of switches \((M = 27.01, \text{S.D.} = 10.00)\) was correlated with psychomotor poverty, \( r(34) = 0.00, p = 0.50; r(34) = 0.02, p = 0.45 \), respectively. However, the correlation between number of clusters \((M = 6.64, \text{S.D.} = 3.20)\) and psychomotor poverty bordered on significance, \( r(34) = -0.27, p = 0.06 \). This correlation became highly significant when total number of words produced was taken into account, \( r(34) = -0.42, p = 0.005 \). For number of switches, the raw score was significantly correlated with the disorganization syndrome, \( r(34) = -0.28, p = 0.05 \). Reality distortion was not significantly correlated with switching, clustering, or total score.

The impact of possible mediator variables was also explored. Age, sex, education, length of illness and
SSPI total score were not significantly correlated with any of the dependent measures, with the exception of the following: (1) age correlated significantly with incongruent Stroop RT, but not with Stroop interference, on the single-trial Stroop test, \( r(34) = 0.35, p = 0.02 \), \( r(34) = 0.21, p = 0.11 \), respectively; and (2) education correlated significantly with number of clusters on verbal fluency, \( r(34) = 0.33, p = 0.03 \) (\( r = 0.38 \) when divided by total number of words). The former association does not alter the current set of results, because Stroop speed did not correlate with the symptom profiles; moreover, because Stroop interference was not correlated with age, the correlation of raw Stroop speed with age can be attributable to color naming speed, not to interference from the conflicting word. Finally, because education is uncorrelated with psychomotor poverty, \( r(34) = 0.00, p = 0.50 \), education could not function as a mediator of the association between psychomotor poverty and number of clusters on verbal fluency.

**4. Discussion**

The purpose of this study was to evaluate whether inconsistencies in the literature investigating associations between cognitive performance (viz., Stroop and verbal fluency) and the symptoms of schizophrenia may be due to variations in test format or failures in reporting all possible test parameters. For both the computerized and list versions of the Stroop test, only errors, not speed, correlated with disorganization. For verbal fluency, a reduction in clustering, not switching, was associated with psychomotor poverty, particularly when clustering was divided by the total number of words produced.

This set of results suggests that inconsistencies in the Stroop literature cannot be explained by differences in test form, but may be attributable to incomplete reporting of parameters (i.e., speed vs. accuracy). Impairment on the Stroop task may manifest as either RT slowing or increased errors, depending upon factors such as the approach of the subject, and/or the emphasis imposed by task instructions. A focus on speed may lead to the manifestation of impairment as an increase in errors, whereas a focus on accuracy may lead to manifestation of impairment as a decrease in speed. Either speed or accuracy should be considered valid measures of susceptibility to Stroop interference, and both should be collected and reported in order to avoid inconsistencies in the literature. Neurobiological research suggests that the anterior cingulate is critically involved in both Stroop interference, (Carter et al., 2000; MacDonald et al., 2000; Peterson et al., 1999; Ruff et al., 2001) and the disorganization syndrome (Liddle, 2001; Liddle et al., 1992; McGuire et al., 1998). Disorganized patients would, therefore, be expected to demonstrate impairment in the selection of cognitive activity due to anterior cingulate dysfunction, and apparently this may manifest and either RT slowing or increased errors.

We observed that for verbal fluency, clustering, but not switching, was associated with psychomotor poverty. The absence of an association between switching mental set and psychomotor poverty is consistent with previous work reporting an absence of an association between psychomotor poverty and tests assessing mental flexibility, such as the Wisconsin Card Sorting Test (Liddle and Morris, 1991; Moritz et al., 2001a; Morrison-Stewart et al., 1992; O’Leary et al., 2000). Thus, impairment in elaboration on a selected mental set (as measured by clustering), not switching between mental sets (as measured by switching), may underlie psychomotor poverty. This is congruent with previous accounts stating that an impaired ability to self-initiate mental activity underlies psychomotor poverty (Liddle, 2001), although in this account, the emphasis was not placed specifically on the concept of elaboration. Recent evidence suggests that both left and right frontal lobe dysfunction are associated with psychomotor poverty (Ebmeier et al., 1993; Liddle et al., 1992; Merrin and Floyd, 1996; Wolkin et al., 1992), leading to the proposal that the impaired ability to elaborate on mental activity associated with psychomotor poverty may be attributable to frontal dysfunction.

The observed correlation between switching (in verbal fluency) and the disorganization syndrome is consistent with prior reports of a significant negative relationship between the disorganization syndrome and tests assessing mental flexibility, such as Trails B and the Wisconsin Card Sorting Test (Liddle and Morris, 1991; Moritz et al., 2001a). This also serves as a partial explanation for previously reported associations between verbal fluency total score and the
disorganization syndrome (Frith et al., 1991; Himelhoch et al., 1996; Liddle and Morris, 1991; Moritz et al., 2001a). This association is congruent with an account holding that dysfunction of the anterior cingulate is associated with disorganization, and results in impaired selection of appropriate cognitive activity.

Using the SSPI, we replicated the three-factor structure reported previously (Arndt et al., 1991; Bilder et al., 1985; Liddle, 1987). The rather high loading of delusions on the disorganization factor may be attributable to some patients’ expansive description of bizarre delusions. Intertwining of the reality distortion and disorganization syndromes has been reported previously (e.g., Bilder et al., 1985). This research was carried out in a long-term care institution; therefore, bizarre delusions may be more common than in samples from short-term care institution. The present study showed, in a direct comparison, that parameters derived from standard neuropsychological tests differ with respect to their association with the syndromes of schizophrenia. It follows that assessing and reporting all parameters can help to eliminate between-study inconsistencies that are due to these factors. Future studies which attempt to identify other sources of between-study inconsistencies in associations between cognitive tests and the symptoms of schizophrenia may focus on the effect of syndrome score construction method, medication type (e.g., see Himelhoch et al., 1996), and the population under study (e.g., chronic vs. short-stay care facilities).

Acknowledgements

The authors thank Jennifer Whitman, Tonya Kragelj, Kimberly Stringer and Lena Freeman for their assistance with the data collection and data management; Dr. A. Troyer for her guidance on verbal fluency scoring; and Dr. K Christensen for her comments on previous versions of this manuscript. This research was supported by a Postdoctoral Fellowship from the Canadian Psychiatric Research Foundation and the Canadian Institute for Health Research to TSW, an operating grant from the Dr. Norma Calder Schizophrenia Foundation to TSW, an operating grant from the Riverview Hospital Academic Steering Committee to TSW and SM and CCR.

References


