Easier Tasks Can Have Better Discriminating Power: The Case of Verbal Fluency

Meredith R. D. Melinder, Deanna M. Barch, Gitry Heydebrand, and John G. Csernansky
Washington University

Loren and Jean Chapman highlighted the importance that the psychometric characteristics of a task may have when comparing different populations (L. J. Chapman & J. P. Chapman, 1973, 1978). Specifically, they pointed out that when comparing 2 tasks measuring different constructs, there is a greater likelihood of finding larger deficits on a task with higher discriminating power, irrespective of what the task measures. While researchers have addressed this issue, they typically have assumed that more difficult tasks have greater discriminating power, often without actually measuring discriminating power. In this article, the authors present data from 2 studies using phonologic and category fluency tasks to demonstrate that the critical factor for task selection and matching is discriminating power, not task difficulty.

Keywords: Discriminating power, verbal fluency, schizophrenia, differential deficit

In their seminal work in the 1970s Loren and Jean Chapman highlighted the importance of examining the psychometric characteristics of tasks used to compare different populations, including factors such as task difficulty and discriminating power (Chapman & Chapman, 1973, 1978). The Chapmans defined discriminating power as “the extent to which the score differentiates the more able from the less able subjects and, hence, differentiates two groups that differ in the ability measured by the test” (Chapman & Chapman, 1973, p. 380). Specifically, they raised the issue that, when comparing two tasks that putatively measure different constructs, there is a greater likelihood of finding larger deficits on the task that has higher discriminating power, irrespective of what the task measures. Since the Chapmans brought this issue to the forefront, researchers have recognized that studies attempting to identify differential deficits in particular cognitive functions in individuals with schizophrenia need to use tasks matched on discriminating power (Strauss, 2001). Most discussions of this problem have used examples in which the more difficult task had greater discriminating power. Therefore, researchers have often assumed that larger group differences on “easier” tasks compared with “harder” tasks can be straightforwardly interpreted as evidence of a differential deficit, without directly examining the relative discriminating power of the tasks (Bokat & Goldberg, 2003; Gourovitch, Goldberg, & Weinberger, 1996; Kremen, Seidman, Faraone, & Tsuang, 2003).

The goal of the current paper is to reemphasize that the critical factor for task matching is discriminating power, not task difficulty, and that an easier task can sometimes have higher discriminating power if it has better reliability and true-score variance (Chapman & Chapman, 1973, 1978). To illustrate this point, we provide data and analyses from two studies of phonologic and category fluency in individuals with schizophrenia.

Verbal Fluency

Verbal fluency measures have been used frequently in the neuropsychological literature as a means of estimating semantic system functioning by comparing performance on phonologic and category versions of the task. Phonologic fluency tasks are thought to depend on the ability to use phonologic cues as a means of accessing lexical representations, and are not thought to require the processing of semantic information. In contrast, category fluency tasks are thought to be dependent on the ability to use semantic representations to generate items, and good performance on such tasks is thought to be dependent on an intact semantic system. The rationale behind comparing phonologic and category fluency as a way of examining semantic processing has been that both tasks theoretically require the same general cognitive processes, with the exception that the latter task requires access to the semantic system (Goldberg et al., 1998). In other words, both tasks theoretically require attention, speed, strategic planning, working memory, and so forth, but only category fluency requires that the participant search through his or her memory based on the semantic aspects of words. Thus, by comparing an individual’s performance on category and phonologic fluency, one can see how the semantic system is functioning, over and above other cognitive functions inherent in the tasks.

Most studies of fluency consistently find that individuals with schizophrenia perform worse than healthy controls on both phonologic and category fluency measures (Allen, Liddle, & Frith, 1993; Bokat & Goldberg, 2003; Bowie et al., 2004; Elvevag,
Discriminating Power

Previous studies of verbal fluency in schizophrenia have not addressed the issue of whether phonologic and category fluency tasks are matched on discriminating power. One reason this has not been addressed may be the belief that it was not necessary because category fluency tasks are typically easier for healthy controls than are phonologic fluency tasks; that is, healthy controls produce more items per category than per letter. Thus the interpretation of such studies relies on the assumption that category fluency tasks have less discriminating power because they are easier, so that greater deficits on category fluency tasks compared with phonologic fluency tasks among individuals with schizophrenia cannot be due to a generalized deficit. However, discriminating power is essentially true-score variance (the component of variance that is replicable and not due to error or factors unrelated to the construct of interest). In turn, true-score variance is the product of task reliability and observed-score variance (the sum of true-score and error variance) (Chapman & Chapman, 1973, 1978).

Task difficulty can greatly influence both reliability and observed-score variance in that more difficult tasks often have greater observed-score variance than do easier tasks. However, more difficult tasks do not always have greater reliability or greater observed-score variance. Thus task difficulty cannot be equated with discriminating power. In fact, the Chapmans and colleagues (Miller, Chapman, Chapman, & Collins, 1995) provided a number of examples in which an easier task with binary items (e.g., scored 0 or 1) had greater true-score variance than a harder task (e.g., accuracy of approximately 50% compared with approximately 20%). The situation is further complicated in tasks in which items are not binary, but instead reflect some sort of count that often has a large range. For example, the Chapmans and others have demonstrated that in reaction-time data (which can be thought of as analogous to count data), there is often a linear relationship between mean latencies and standard deviations, such that slower participants (i.e., higher reaction time counts) have larger standard deviations (i.e., observed-score variance) (Chapman, Chapman, Curran, & Miller, 1994; Faust, Balota, Spieler, & Ferraro, 1999; Hale, Myerson, Smith, & Poon, 1988). Individuals typically produce more items for category fluency tasks than for phonologic fluency tasks (e.g., higher counts). If phonologic and category fluency tasks have similar reliabilities, but category fluency has greater observed-score variance, then category fluency tasks would have greater true-score variance and discriminating power. If so, then one cannot rule out the possibility that the relatively greater deficits that individuals with schizophrenia show on category fluency tasks compared with phonologic fluency tasks reflect a generalized deficit rather than a deficit in some specific cognitive process tapped by the category fluency task.

Few of the previous studies comparing phonologic and category fluency in schizophrenia have provided data suggesting comparable reliability for the two types of tasks. However, examination of the means and standard deviations for the control groups in previous studies does suggest that observed-score variance is generally higher for category fluency compared with phonologic fluency. For example, Elvevag et al. (2001) found greater variance for category fluency compared with phonologic fluency in healthy controls, as well as a differential deficit among individuals with schizophrenia on category fluency tasks. Similarly, Rossell et al. (1999) found greater variance for category fluency compared with phonologic fluency, and they found a differential deficit in category compared with phonologic fluency among deluded patients. Of note, not all studies that show greater variance in category fluency versus phonologic fluency among healthy controls have found differential deficits in category fluency among individuals with schizophrenia (Joyce et al., 1996; Zalla et al., 2004). However, because these studies did not provide reliability data, the discriminating power of category fluency and phonologic fluency could not be compared across studies.

The goal of the current project was to directly compare the discriminating power for category fluency and phonologic fluency tasks and to evaluate the relationship of discriminating power to group difference effect sizes in individuals with schizophrenia and demographically similar controls. Specifically, we analyzed data from two different studies (i.e., Study 1 and Study 2) that administered versions of both category and phonologic fluency tasks. We show that when analyzed in the typical fashion, with phonologic and category fluency unmatched for discriminating power, category fluency elicits significantly larger group differences between individuals with schizophrenia and healthy controls than does phonologic fluency. However, when we created dependent variables for the category fluency and phonologic fluency tasks that were matched on discriminating power, there were equal deficits among individuals with schizophrenia in phonologic fluency and category fluency.

Study 1

Method

Participants

We recruited 55 outpatients diagnosed with schizophrenia or schizoaffective disorder from local outpatient mental health treatment facilities to participate in this study. Diagnoses were based on the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders.
Tasks

Verbal fluency. For phonologic fluency, participants were asked to name as many words beginning with a certain letter of the alphabet as they could within the time limit. They were told to exclude proper nouns and different forms of the same word. Each participant completed this task for three phonologic categories: the letters F, A, and S. For category fluency, participants were asked to name as many exemplars of a category as they could within the time limit. Each participant completed this task for the three semantic categories: animals, fruits, and body parts. Fifty participants (28 healthy controls and 22 individuals with schizophrenia) were asked to name words within a 120-s time limit, and their responses in the first 60 s were used for this study. Forty-two participants (17 healthy controls and 25 individuals with schizophrenia) were asked only to name words within a 60-s time limit. The performance within the first 60 s for participants who completed the 120-s task was not significantly different from that of the participants who completed only the 60-s task. We collected test–retest reliability data for each trial of both the phonologic and category fluency tasks used in Study 1 and in Study 2 (60-s versions) from 21 additional controls who were demographically similar in age and parental education to the controls used in this study and in Study 2. The time interval between Test 1 and Test 2 was either 1 or 2 days.

WRAT–III Reading. The Reading subtest of the WRAT–III, which requires that participants read a list of words increasing in difficulty, was administered as an estimate of premorbid IQ. The basis for administering this test as a measure of premorbid IQ is the finding that reading abilities are often preserved in psychotic disorders even when other cognitive functions have declined (Dalby & Williams, 1986). The raw scores on the Reading subtest were converted to age-scaled standard scores on the basis of the test guidelines (Wilkinson, 1993).

Results

The demographic and clinical ratings of both participant groups are shown in Table 1. Individuals with schizophrenia and healthy controls did not differ significantly in age, 

\[
t(90) = -0.023, \quad p > .98; \text{ level of parental education, } t(90) = 1.106, \quad p > .27; \text{ gender, } \chi^2(2, N = 92) = 0.03, \quad p > .86; \text{ or race, } \chi^2(2, N = 92) = 1.21, \quad p > .55.
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Healthy controls (n = 45)</th>
<th>Participants with schizophrenia (n = 47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>M = 37.89, SD = 9.15</td>
<td>M = 37.94, SD = 10.26</td>
</tr>
<tr>
<td>Sex (% male)</td>
<td>57.8</td>
<td>59.6</td>
</tr>
<tr>
<td>Race (% Caucasian)</td>
<td>62.2</td>
<td>59.6</td>
</tr>
<tr>
<td>Education (years)</td>
<td>16.47*</td>
<td>12.85*</td>
</tr>
<tr>
<td>Parental education (years)</td>
<td>14.31</td>
<td>13.74</td>
</tr>
<tr>
<td>Age at first hospitalization (years)</td>
<td>8.15</td>
<td>8.27</td>
</tr>
<tr>
<td>WRAT–III Reading standard score</td>
<td>106.40*</td>
<td>97.79*</td>
</tr>
</tbody>
</table>

* \( p < .001 \).
subtrials within each type of fluency task) for the control participants using Cronbach’s alpha. We measured test–retest reliability for each individual category and phonologic fluency item, as well as the sum scores, using correlations computed on the additional test–retest sample of controls. The internal consistency of the two verbal fluency tasks was relatively similar: .78 for phonologic fluency and .70 for category fluency. In addition, the test–retest reliability for the sum of the three trials for each verbal fluency task was high and very similar. The test–retest reliability of the individual trials varied somewhat within each task, but all trials had moderate to good test–retest reliability. We then calculated the true-score variance (Test–Retest Reliability × Observed-Score Variance) for a number of different ways of computing a dependent variable, including: (a) the average across all three trials within each task, (b) the summed total across all three trials, and (c) each individual trial of each task.

As can be seen in Table 2, although the verbal fluency tasks were similar on reliability, the observed-score variance for phonologic and category fluency was different. When we calculated discriminating power (Reliability × Observed-Score Variance) for either the total scores (sum of the three trials for each type) or the average (average of the three trials for each type), category fluency had nearly twice the true-score variance of phonologic fluency. Thus, category fluency has greater discriminating power than phonologic fluency. As such, it allows us to predict a priori that individuals with schizophrenia will show a larger deficit on the category fluency task than on the phonologic fluency task.

To test the hypothesis that this difference in discriminating power might account for observed differences between performance on the different fluency tasks in the participants with schizophrenia, we first conducted a group comparison without taking into account differences in discriminating power. We analyzed the data using a three-factor analysis of variance (ANOVA), with diagnostic group (patient, control) as a between-subjects factor and both fluency task type (category, phonologic) and trial (F, A, S for phonologic; animals, fruits, body parts for category) as within-subject factors. As expected, there was a significant main effect for group, $F(1, 90) = 33.3, p < .001$, and a significant main effect for fluency task type, $F(1, 90) = 98.5, p < .001$. In addition, as predicted, there was a significant diagnostic Group × Fluency Task Type interaction, $F(1, 90) = 14.7, p < .001$. As shown in Table 2, there were larger group differences for category compared with phonologic fluency. There were no significant group differences in the number of intrusions or perseverations during the phonologic or category fluency tasks. We also analyzed the data using a two-factor ANOVA with total scores (sum of the three trials for each fluency task type) as the dependent variable, diagnostic group as a between-subjects factor, and fluency task type as a within-subject factor. This analysis again revealed a main effect of group, $F(1, 90) = 32.8, p < .001$; a main effect of fluency task type $F(1, 90) = 89.1, p < .001$; and a Group × Fluency Task Type interaction, $F(1, 90) = 17.1, p < .001$ ($\eta^2 = .16$). As with the previous analyses, this interaction reflected the fact that there were larger group differences for category as compared with phonologic fluency.

If the only reason for this “differential deficit” on category versus phonologic fluency is the psychometric differences between the tasks, we should be able to eliminate this greater deficit on category fluency by matching the fluency tasks on discriminating power. On the other hand, if this difference truly reflects a differential deficit on semantic processing as indexed by category fluency deficits, we should find that a Group × Fluency Task Type interaction remains, even when the tasks are matched for discriminating power. To develop dependent measures for category and phonologic fluency tasks that were matched for discriminating power, we examined the true-score variance for individual trials of the category and phonologic fluency tasks, as well as combinations of trials. The only measure of reliability that we had for the individual trials was test–retest reliability. Thus, we computed true-score variance as the product of test–retest reliability and observed-score variance for each trial. For example, as shown in Table 2, the letter S trial from phonologic fluency and the fruit trial

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Healthy controls (n = 45)</th>
<th>Participants with schizophrenia (n = 47)</th>
<th>Test–retest reliability (controls)</th>
<th>Variance (controls)</th>
<th>True-score variance (controls)</th>
<th>Group difference effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totala</td>
<td>42.2</td>
<td>9.8</td>
<td>34.8</td>
<td>12.5</td>
<td>.81</td>
<td>96.3</td>
</tr>
<tr>
<td>Averagea</td>
<td>14.1</td>
<td>3.3</td>
<td>11.5</td>
<td>4.2</td>
<td>.81</td>
<td>10.7</td>
</tr>
<tr>
<td>F</td>
<td>14.6</td>
<td>3.7</td>
<td>12.0</td>
<td>4.9</td>
<td>.64</td>
<td>13.3</td>
</tr>
<tr>
<td>A</td>
<td>12.8</td>
<td>4.0</td>
<td>9.6</td>
<td>4.4</td>
<td>.51</td>
<td>15.9</td>
</tr>
<tr>
<td>S</td>
<td>14.8</td>
<td>4.1</td>
<td>13.0</td>
<td>5.0</td>
<td>.85</td>
<td>16.8</td>
</tr>
<tr>
<td>Letters A + S</td>
<td>27.6</td>
<td>7.1</td>
<td>22.6</td>
<td>8.5</td>
<td>.76</td>
<td>50.1</td>
</tr>
<tr>
<td>Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totalb</td>
<td>60.4</td>
<td>13.5</td>
<td>41.9</td>
<td>14.1</td>
<td>.85</td>
<td>183.0</td>
</tr>
<tr>
<td>Averageb</td>
<td>20.1</td>
<td>4.5</td>
<td>14.2</td>
<td>4.3</td>
<td>.85</td>
<td>20.3</td>
</tr>
<tr>
<td>Animals</td>
<td>20.6</td>
<td>5.7</td>
<td>14.6</td>
<td>5.0</td>
<td>.74</td>
<td>32.8</td>
</tr>
<tr>
<td>Fruits</td>
<td>15.4</td>
<td>3.9</td>
<td>11.3</td>
<td>3.7</td>
<td>.78</td>
<td>14.9</td>
</tr>
<tr>
<td>Body parts</td>
<td>24.4</td>
<td>7.1</td>
<td>16.9</td>
<td>5.8</td>
<td>.84</td>
<td>50.0</td>
</tr>
</tbody>
</table>

a Internal consistency = .78. b Internal consistency = .70.
from category fluency were the most similar in discriminating power. In addition, the sum of the letters A and S trials for phonologic fluency and the animals trial for category fluency were also similar in discriminating power.

To then examine whether the Group × Fluency Task Type interaction remained significant when using these matched dependent variables, we conducted a two-factor ANOVA with diagnostic group as a between-subjects factor and fluency task type as a within-subject factor, using only the letter S trial for phonologic fluency and the fruit trial for category fluency. We once again found a significant main effect of group, $F(1, 90) = 7.0, p < .01$. However, there was no longer a significant Group × Fluency Task Type interaction, $F(1, 90) = 0.6, p > .40$ ($\eta^2 = .04$). We found similar results in a two-factor ANOVA with group as a between-subjects factor and fluency task type as a within-subject factor, using the sum of the A and S trials for phonologic fluency and the animals trial for category fluency. Specifically, there was a significant main effect of group, $F(1, 90) = 21.5, p < .001$, but no significant Group × Fluency Task Type interaction, $F(1, 90) = 0.5, p > .5$, ($\eta^2 = .005$).

Discussion

The results of this study suggest that prior findings of differential deficits on category fluency as compared with phonologic fluency tasks among individuals with schizophrenia may reflect, at least in part, psychometric confounds. Specifically, when we conducted analyses using the same types of dependent variables as used in prior studies (Feinstein et al., 1998; Goldberg et al., 1998; Gourvitch et al., 1996; Kremen et al., 2003; Rossell et al., 1999), we replicated previous findings of a differential deficit on category compared with phonologic fluency. However, analyses of discriminating power also indicated that when analyzed in the traditional way, category fluency had greater discriminating power than phonologic fluency. Therefore, greater impairment on category compared with phonologic fluency could have reflected a generalized deficit. When we used dependent variables for category and phonologic fluency that were matched on discriminating power, we no longer found a significant differential deficit on category fluency. Nonetheless, it was still the case that the category fluency variables tended to have larger group difference effect sizes than the phonologic fluency variables that were similar on discriminating power (see Table 2). However, the effect sizes of the Group × Fluency Task Type interactions were much smaller for variables similar on discriminating power ($\eta^2 = .04$ and .005) than the effect sizes for the original analyses ($\eta^2 = .16$). Taken together, these results indicate that analyses of fluency variables matched on discriminating power provide less support for the hypothesis that individuals with schizophrenia have a differential deficit in semantic processing than do analyses in which the category fluency variables have higher discriminating power than the phonologic fluency variables. Further, the results highlight the fact that easier tasks can sometimes have better discriminating power than harder tasks.

Study 2

The goal of Study 2 was to determine whether we could replicate the results of Study 1 using a separate sample of individuals diagnosed with schizophrenia and healthy controls.

Method

Participants

Participants included 53 individuals with DSM–IV–TR schizophrenia and 60 healthy controls between the ages of 18 and 59. The individuals with schizophrenia were all clinically stable inpatients and outpatients recruited through an ongoing structural imaging protocol conducted by John G. Csernansky. Healthy controls were recruited from the same community as the individuals with schizophrenia through local advertisements. Healthy controls were excluded if they had any lifetime history of Axis I psychiatric disorder or any first-order family member with a psychotic disorder. Potential participants (either individuals with schizophrenia or healthy controls) were also excluded for the following reasons: (a) meeting DSM–IV (American Psychiatric Association, 1994) criteria for substance abuse (severe) or dependence (any type) at any time within the past 3 months; (b) the presence of any clinically unstable or severe medical disorder, or a medical disorder that would confound the assessment of psychiatric diagnosis or make participation in the research protocol unsafe; (c) present or past head injury with documented neurologic sequelae or that caused loss of consciousness; or (d) meeting DSM–IV criteria for mental retardation (mild or greater in severity). Parental socioeconomic status was measured using the Hollingshead Index (Hollingshead & Redlich, 1958). All individuals with schizophrenia were medicated, taking a variety of typical and atypical antipsychotic medications.

Schizophrenia and control diagnoses were determined using the SCID-IV (First et al., 2001). The structured interviews were conducted by a master’s in social work-level research assistant who had completed SCID-IV training and regularly participated in ongoing diagnostic training sessions at the Metropolitan Psychiatric Center. The SCID-IV interviewer had access to all data from present and past Metropolitan Psychiatric Center hospital records, corroborative personal sources (e.g., family), and records from other hospitals. In addition, a semistructured interview was performed by an expert clinician (in most cases, John G. Csernansky), also using DSM–IV criteria. This expert clinician also had access to all available medical records and collaborative sources but was blind to the results of the SCID-IV interview. The participant’s final diagnosis was determined by a consensus meeting between the SCID-IV interviewer and the expert clinician. Informed consent was obtained from all participants prior to study participation in accordance with the Washington University and Metropolitan Psychiatric Center institutional review boards, and all participants were paid for their participation.

Verbal Fluency Tasks

As part of a larger neuropsychological battery, the participants were administered versions of both category and phonologic fluency tasks. In the phonologic fluency task, participants were asked to name as many words beginning with a certain letter of the alphabet as they could within the time limit. They were told to name words that begin with that letter, excluding proper nouns and different forms of the same word. Each participant completed this task for two phonologic categories: the letters S and P. Likewise, for category fluency, the participants were asked to name as many exemplars of a category as they could within the time limit. Each participant completed this task for the one semantic category, which was animals. All participants were asked to name the words for each trial within a 60-s time limit.

Results

The demographic and clinical characteristics of both participant groups are shown in Table 3. The groups did not differ significantly on age, $t(111) = 0.12, p > .9$; gender, $\chi^2(2, N = 113) = .25, p > .5$; or parental socioeconomic status, $t(111) = -1.0, p > .3$. 
We began the analysis of data from Study 2 by examining the relative discriminating power of the two fluency tasks, using the internal consistency estimates from Study 1. With this type of data set (two trials for fluency, one trial for category), one might choose to analyze the data by comparing the average of the two phonologic fluency trials to the one category fluency trial. As shown in Table 4, the observed-score variance was higher for the category fluency trial compared with the average of the two phonologic fluency trials, leading to greater true-score variance and greater discriminating power. Therefore, one would predict that there would be a larger group difference on the category compared with the phonologic fluency tasks. However, as shown in Table 4, the sum of the two phonologic fluency trials (compared with the average) had a higher observed-score variance. Thus, true-score variance and discriminating power were more similar to the true-score variance and discriminating power for category fluency. Therefore, one no longer would predict a larger group difference on category as compared with phonologic fluency. Again, however, if there were truly a differential deficit in semantic processing among individuals with schizophrenia, one would predict that the larger group differences on category fluency would remain even when using dependent measures matched for discriminating power.

To test these predictions, we first analyzed the data using the average phonologic fluency dependent measure, using a two-factor ANOVA with diagnostic group as a between-subjects factor and fluency task type (phonologic, category) as the within-subject variable. This ANOVA revealed a main effect of diagnostic group, $F(1, 111) = 48.0, p < .001$, and a main effect of fluency task type, $F(1, 111) = 84.5, p < .001$. In addition, as predicted, we found a significant diagnostic Group × Fluency Task Type interaction, $F(1, 111) = 13.0, p < .001, \eta^2 = .11$. We next analyzed the data using the sum of the two phonologic fluency trials as the dependent measure (matched to category fluency on discriminatory power), using a 2-factor ANOVA with diagnostic group as a between-subjects factor and fluency task type (phonologic, category) as the within-subject variable. We again found a significant main effect of group, $F(1, 111) = 40.3, p < .001$, and a significant main effect of fluency task type, $F(1, 111) = 84.5, p < .001$. However, the Group × Fluency Task Type interaction was no longer significant, $F(1, 111) = 0.3, p > .6 (\eta^2 = .002)$.

### Discussion

The goal of the current study was to illustrate that an easier task can have higher discriminating power than a harder task, if the easier task has either higher reliability or greater observed-score variance. We used a comparison of phonologic and category fluency tasks to illustrate this point. As described earlier, a number of previous studies have found Group × Fluency Task Type interactions that have been interpreted as reflecting a differential deficit in category fluency among individuals with schizophrenia. However, our analyses of the relative discriminating power of phonologic and category fluency tasks show that category fluency has higher discriminating power, even though healthy controls typically perform better on category versus phonologic fluency. This higher discriminating power primarily results from greater observed-score variance, as the internal consistency and test–retest reliability of the two tasks are fairly similar.

### General Discussion

Similarly to prior studies, we found significant Group × Fluency Task Type interactions for both Study 1 and Study 2 when the data were analyzed without taking into account differences in discriminating power. However, when we used dependent variables for each fluency task type that were matched for discriminating power, the effect sizes for these interactions were much reduced and the interactions were no longer significant. If individuals with schizophrenia had a differential deficit in semantic processing as indexed by category fluency performance, one would have expected the magnitude of the Group × Fluency Task Type interaction to remain stable even when the dependent variables were matched on discriminating power. Of note, it was the
case that the category fluency variables did tend to have higher group difference effect sizes than the phonologic fluency variables, even though the Group × Fluency Task Type interactions were no longer significant when examining dependent variables similar on discriminating power. Even though these studies had relatively large sample sizes and good power to detect significant interactions, it is possible that even larger sample sizes would reveal greater evidence for a differential deficit in semantic processing. In either case, the analyses presented above demonstrate that at minimum, the magnitude of any such interaction with fluency task type is dramatically reduced when comparing variables unmatched on discriminating power to variables matched on discriminating power.

In summary, the current results highlight the importance of examining relative discriminating power in all studies of cognitive and language function in special populations such as individuals with schizophrenia, even when one finds what appears to be a differential deficit on an easier task. Although the Chapmans (Chapman & Chapman, 1973, 1978) clearly articulated this point in their earlier writings, many researchers have relied too much on the role of task difficulty in establishing differential deficits. However, as pointed out by the Chapmans and by the results of the current study, an easier task can clearly have better discriminating power than a harder task. Therefore, without an explicit analysis of discriminating power in addition to task difficulty, one can reach misleading conclusions about the magnitude or even presence of differential cognitive deficits.

References


Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of differential cognitive deficits. In their earlier writings, many researchers have relied too much on the role of task difficulty in establishing differential deficits. However, as pointed out by the Chapmans and by the results of the current study, an easier task can clearly have better discriminating power than a harder task. Therefore, without an explicit analysis of discriminating power in addition to task difficulty, one can reach misleading conclusions about the magnitude or even presence of differential cognitive deficits.

REFERENCES


Dalby, I. T., & Williams, R. (1986). Preserved reading and spelling ability in psychotic disorders. Psychological Medicine, 16, 171–175.


Received June 18, 2003
Revision received November 1, 2004
Accepted January 6, 2005