Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Brain and Language 92 (2005) 240–250
Brain and Language
www.elsevier.com/locate/b&l

Is problem solving dependent on language?☆

Juliana V. Baldo[a,b,*], Nina F. Dronkers[b,c], David Wilkins[b,d], Carl Ludy[b], Patricia Raskin[a], Jiye Kim[a]

[a] Department of Psychology, Scripps College, Claremont, CA, USA
[b] Center for Aphasia and Related Disorders, VA Northern California Health Care System, Martinez, CA, USA
[c] Departments of Neurology and Linguistics, University of California, Davis, USA
[d] Department of Speech and Communication Studies, San Francisco State University, USA

Accepted 11 June 2004
Available online 24 August 2004

Abstract

There has been a long-standing debate in the fields of philosophy and cognitive science surrounding the relationship of language to cognition, but the exact nature of this relationship is still unclear (Sokolov, 1968/1972). In the current study, we explored the role of language in one aspect of cognition, namely problem solving, by administering the Wisconsin Card Sorting Test (WCST) to stroke patients with varying degrees of language impairment (Experiment 1) and to normal participants under conditions of articulatory suppression (Experiment 2). In Experiment 1, there was a significant correlation between performance on the WCST and language measures such as comprehension and naming. Demonstrating the specificity of this result, we also found a significant relationship between language performance and another test of problem solving, the Raven's Colored Progressive Matrices, but no relationship between language and a test of visuospatial functioning. In Experiment 2, normal participants were significantly impaired on the WCST under conditions of articulatory suppression, relative to a baseline condition. Together, these findings suggest that language plays a role in complex problem solving, possibly through covert language processes.

Keywords: Aphasia; Wisconsin Card Sorting Test; Problem solving; Executive functions; Reasoning; Language

1. Introduction

Problem solving generally requires a series of steps, including analysis of a problem, generating possible solutions, testing solutions, and modifying behavior/switching strategies when a solution is unsuccessful. This ability has most often been associated with prefrontal cortex, as patients with frontal lesions often show impaired problem solving due to impaired strategy formation and an inability to modify behavior based on feedback (Baldo, Delis, Wilkins, & Shimamura, 2004; Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Klouda & Cooper, 1990; Miller & Tippett, 1996; Stuss & Alexander, 2000). Here, we explore the hypothesis that another component may be crucial for normal problem solving: language. By language, we mean a rule-based, symbolic representation system that may be instantiated in a number of different ways (i.e., not just spoken language). Experimental evidence to directly test the relationship between problem solving and language has been somewhat elusive. This issue has been addressed in studies of neuropsychological performance in patients with speech and language deficits (e.g., Basso, De Renzi, Faglioni, Scotti, & Spinnler, 1973;

☆ We would like to thank Joe Elder, Jary Larsen, Brenda Redfern, Kimberly Mae Wong, and Emily Telleen-Lawton for their assistance in collecting data for this study. We acknowledge the technical expertise of Stephen Wilson who developed the lesion mapping software. Finally, we thank You Young Kang for her assistance in recording the rhythm sequence used in Experiment 2.
* Corresponding author. Fax: +1-925-372-2553.
E-mail address: jbaldo@scrippscollege.edu (J.V. Baldo).

0093-934X/$ - see front matter
doi:10.1016/j.bandl.2004.06.103

Hamsher, 1991; Piercy, 1964; Weinstein & Teuber, 1957) and in findings from children's acquisition of linguistic and cognitive abilities (e.g., Gentner, 2003).

Folk theory and subjective experience suggest that we often engage in some form of inner speech and that this on-line discourse facilitates our ability to think through problems (Carruthers, 2002; Hurlburt, 1990; Sokolov, 1968/1972). Sokolov reviewed a number of studies in which tasks such as mental arithmetic and word repetition were shown to be accompanied by muscular contractions in the articulators (i.e., tongue, lips, etc.). Such findings would suggest that thought is sometimes accompanied by covert vocalization or "inner speech" (see Sokolov). Recent evidence suggests that such internal verbalizations do indeed support normal cognition. Hermer-Vazquez, Spelke, and Katsnelson (1999) used a dual task paradigm and found that verbal shadowing (repeating tape-recorded speech) disrupted participants' ability to perform a spatial-localization task. Interestingly, they found that the way participants solved the task under these conditions was similar to the way in which young children and other animals do. They concluded that language, specifically covert verbalization, normally supports flexible thinking and adult human cognition.

A number of older studies in the literature suggested a relationship between language and problem solving by testing patients with aphasia on standardized tests of intelligence and problem solving (Archibald, Wepman, & Jones, 1967; Borod, Carper, & Goodglass, 1982; De Renzi, Faglioni, Savoiardo, & Vignolo, 1966; Edwards, Ellams, & Thompson, 1976; Hjelmquist, 1989; Larrabee & Haley, 1986; but see Basso et al., 1973). Some of these papers reported that cognitive impairment was most striking in patients with severe comprehension deficits, such as patients with Wernicke's aphasia (e.g., Hjelmquist; Kertesz & McCabe, 1975). Such findings imply that language supports cognition. However, other researchers have argued that such relationships are instead due to a coincidence of anatomy, that is, "intelligence deficits" in aphasic patients are due to the encroachment of lesions on nearby regions crucial for cognition (see Basso et al., 1973; Hamsher, 1991).

In the current study, we were interested in further exploring the idea that problem solving directly relies on language. We did this by testing problem solving in patients with varying degrees of language impairment (Experiment 1) and in normal participants whose inner speech was disrupted by experimental manipulation (Experiment 2). In the two experiments presented here, we tested participants on the Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtis, 1993), which is probably the most recognized and well-studied test of problem solving used today. It requires the examinee to sort cards into different piles, based on three different salient features: color, shape, and number. The examiner simply responds "correct/incorrect," and the examinee must learn how to sort the cards based on this feedback. Periodically, the examiner switches the sorting criterion category (without alerting the examinee), and the examinee must learn the new sorting rule based on feedback. This task, like many tests of problem solving, requires the examinee to keep track of responses and feedback, to update information in working memory accordingly, and to switch strategies when one fails.

The WCST seemed ideal for testing the relationship between language and problem solving because it involves minimal instructions and does not require any overt verbalization from the examinee. In this way, we could test the hypothesis that language supports problem solving by testing patients with a range of language impairments, avoiding the potential confound of overt response deficits such as are needed in other problem solving tasks. At the same time, the shapes, colors, and number of items represented on the stimulus cards are all easily named (e.g. "two blue crosses"). As Baddeley (2000) observed, "adult subjects typically opt to name and subvocally rehearse visually presented items, thereby transferring the information from a visual to an auditory code" (p. 419). Thus, the current study assessed the relationship between language processing and complex problem solving.

Although some studies have suggested a link between problem solving and language, a number of questions remain: Is problem solving dependent on language in the normal individual? Which aspects of speech and language (e.g., fluency, comprehension, etc.) are most crucial for problem solving? Which aspects of problem solving are most dependent on language? How specific are problem-solving deficits in aphasic patients? Do specific brain regions correlate with impaired problem solving in patients with aphasia? We sought to address these questions in the current study.

2. Experiment 1: WCST in patients with language impairment

In Experiment 1, we tested a large series of stroke patients on a standardized task of problem solving, the Wisconsin Card Sorting Task. Unlike earlier studies of aphasia, in the current study, we were able to conduct extensive language and cognitive testing with patients, as well as reconstruct and localize their lesions based on three-dimensional MRI and CT scans. In this way, we were able to address the role of language in problem solving performance, as well as analyze the contribution of different brain regions to this ability. We were also interested in determining which aspects of problem solving would be most problematic for patients with speech and language deficits. Based on our clinical observations and earlier studies, we predicted that comprehension deficits would be most predictive of WCST performance and that flexibility on the WCST would be most disrupted by language impairment.

2.1. Method

2.1.1. Participants

Participants were 41 stroke patients who were assessed at the Center for Aphasia and Related Disorders at the Veterans Affairs Northern California Health Care System (VANCHCS) in Martinez, CA. Inclusion criteria for the current study included English as a first language, right-handed according to a modified Edinburgh Handedness Inventory, single cerebrovascular accident (CVA; e.g., no tumor, progressive cases, or multiple CVAs), and no history of prior neurologic injury, psychiatric disability, or substance abuse. These strict criteria resulted in the exclusion of 30 additional patients who were originally tested in a larger study with different selection criteria. The resultant group consisted of 41 patients (30 men and 11 women). The mean age of patients was 62.5 years ($SD = 11.3$; range 31–83), mean education was 14.2 years ($SD = 2.8$; range 7–20), and mean months post-onset at time of testing was 51.0 ($SD = 55.2$; range 12–196).

Patients' lesions were reconstructed from CT and MRI scans that were obtained at least 3 weeks post-CVA. Thirty-five of the patients had single left hemisphere CVAs and six patients had single right hemisphere CVAs.

Patients were tested in the chronic phase of their illness, at least 1 year post-onset of the stroke. Diagnosis of aphasia, as well as aphasia classification, was assessed with the Western Aphasia Battery (WAB; Kertesz, 1982). Aphasia classifications included Broca's ($n = 5$); Wernicke's ($n = 4$); conduction aphasia ($n = 4$); anomic aphasia ($n = 4$); transcortical motor ($n = 1$); unclassifiable ($n = 1$); and within normal limits (WNL; $n = 22$). The WNL group included both left and right hemisphere patients, some of whom were diagnosed with aphasia acutely, but at the time of testing were WNL on the WAB (aphasia quotient > 93.7). Many of the left hemisphere WNL patients still had clinically significant speech and language symptoms (e.g., dysfluency and anomia) to which the WAB was not sensitive.

The study was approved by the Institutional Review Board at the VANCHCS. All participants were consented prior to the study according to institutional guidelines.

2.1.2. Materials

Patients were administered the manual version of the Wisconsin Card Sorting Task (WCST; Heaton et al., 1993) and the Western Aphasia Battery (Kertesz, 1982). The WCST consists of 4 key cards and 128 test cards that have different arrays of 1–4 items (stars, triangles, circles, or crosses) drawn in different colors (red, blue, green, and yellow). The WAB is a standardized battery that tests speech, language, and other cognitive functions using paper and pencil tests. The Raven's Colored Progressive Matrices (RCPM; Raven, 1962) and a version of the Block Design test were also administered, as part of the standard WAB administration. On the RCPM, the participant must choose one of six potential designs to complete a given pattern. There are 36 problems, presented on separate pages in a stimulus booklet. The Block Design test requires participants to arrange four red and white blocks to match three different red and white designs.

2.1.3. Procedure

Patients were tested individually in a noise-attenuated testing room. Generally, the WCST and WAB were given in separate testing sessions, separated by a week or more. In most cases, procedures followed standard test instructions as provided in the individual test manuals. Due to comprehension deficits in a few patients, it was sometimes necessary to repeat instructions and/or use gestures to ensure that the task instructions were clear. On the WCST, only one patient failed to sort even one category, which might indicate a failure to comprehend the directions. However, this seems unlikely as the patient thoughtfully placed one card at a time in different piles and responded in appropriate ways to feedback (e.g., saying "OK" when the examiner said "correct," and "darn" when the examiner said "incorrect").

The WCST data were analyzed with the WCST computerized scoring program, which generates both raw scores and a number of age- and education-corrected percentile scores based on normative data. Here, we focus on a subset of the measures that were generated using this automatic scoring program: number of categories sorted (out of 6), number of trials to complete the first category (minimum is 10), correct trials (i.e., trials in which the response matches the sorting principle in effect), perseverative errors (i.e., incorrect trials in which the participant persists in responding to an incorrect dimension), non-perseverative errors (i.e., incorrect trials that are not perseverative), and conceptual level responses (i.e., runs of three or more consecutive correct responses, reflecting insight into the sorting strategy).

From the WAB, we chose to focus on five critical measures of speech and language: overall speech/language ability (aphasia quotient or AQ), fluency, repetition, naming, and comprehension. The AQ score is a summary score based on overall fluency, comprehension, repetition, and naming. The fluency measure is based on patients' spontaneous speech and verbal description of a picture. The repetition measure is based on patients' ability to repeat single words, phrases, and sentences. The naming score reflects performance on naming common objects. Last, the comprehension measure reflects performance on word recognition, yes/no questions, and sequential commands.

The RCPM and Block Design test were also administered as part of the standard WAB administration. The RCPM had 36 trials, and there was no time limit on this test. Patients were instructed (both verbally and with gesture) to choose one of six designs that best completed the pattern in the upper half of the page. The Block Design Test had three trials following a practice trial, and scores were based on speed and accuracy. Total correct scores on both of these tests were calculated for each participant based on standardized manual procedures (Kertesz, 1982).

2.2. Results and discussion

First, we calculated descriptive statistics for WCST performance in the patient group as a whole, including a comparison with normative data (Heaton et al., 1993). The mean number of categories sorted by patients on the WCST was 3.4 ($SD = 2.0$; range 0–6). On average, patients took 21.0 trials ($SD = 21.3$) to complete the first category (range 10–128). Using age- and education-corrected norm tables (Heaton et al., 1993), patients' overall performance as measured by Percent Errors averaged the 27th percentile (range 1st–91st percentile, impaired to high average range). Patients' mean Percent of Perseverative Errors was at the 30th percentile (range 1st–93rd), and mean Percent of Non-Perseverative Errors was at the 29th percentile (range 1st–99th). Last, the mean Percent of Conceptual Level Responses was at the 27th percentile (range 1st–94th). As can be seen, there was a wide range of performance across patients, ranging from impaired to high average performance across the different WCST measures. For example, some patients sorted 0 categories and others 6; and some patients completed the first category in 10 trials (minimum number possible) while others took over 100 trials.

To investigate the role that language played in performance on the WCST, we conducted a series of correlations between patients' speech and language performance on the WAB and performance on the WCST. All patients' data (both right and left hemisphere CVAs) were included in the analyses below, so that a wide range of language dysfunction from none to severe could be correlated with WCST performance. Correlation coefficients were computed (Pearson product–moment correlation, two-tailed) to measure the relationship between speech/language variables (overall ability or AQ, Fluency, Comprehension, Naming, and Repetition) and a number of variables on the WCST (Percent Correct, Perseverative and Non-Perseverative Errors, Conceptual Level Responses, and the number of trials to complete the first category). For these tests, we used raw data, rather than norm-corrected percentiles, since we were more interested in the relationship between language and problem solving and less concerned with normative percentile cut-offs. Also, we did not want to distort the raw data by using age- and education-corrected percentiles.

First, we analyzed overall performance by computing the correlation coefficient for the relationship between Percent Correct (the percentage of correct trials) on the WCST and the WAB variables. Significant relationships were found between Percent Correct and AQ, $r(39) = .33$, $p < .05$, between Percent Correct and Naming, $r(39) = .38$, $p < .05$, and between Percent Correct and comprehension, $r(39) = .43$, $p < .01$. There was no significant relationship between Percent Correct on the WCST and Fluency, $r(39) = .20$, $p = .20$, nor between Percent Correct and Repetition, $r(39) = .24$, $p = .13$.

We also assessed the relationship between overall conceptual understanding of the WCST (Percent of Conceptual Level Responses) and the WAB variables. Again, there were significant correlations between Percent of Conceptual Level Responses and AQ, $r(39) = .32$, $p < .05$, between Percent of Conceptual Level Responses and naming, $r(39) = .40$, $p < .01$, and between Percent of Conceptual Level Responses and Comprehension, $r(39) = .46$, $p < .01$. There was no significant relationship between Percent of Conceptual Level Responses and Fluency, $r(39) = .16$, $p = .32$, nor between Percent of Conceptual Level Responses and Repetition, $r(39) = .21$, $p = .18$.

To evaluate the *types* of errors that patients made on the WCST, we computed correlation coefficients for the relationship between WAB variables and the Percent of Perseverative and Non-Perseverative Errors on the WCST. The Percent of Perseverative Errors was significantly correlated with AQ, $r(39) = -.44$, $p < .01$; with Repetition, $r(39) = -.40$, $p < .05$; with naming, $r(39) = -.52$, $p < .001$; and with Comprehension, $r(39) = -.44$, $p < .01$; but not with Fluency, $r(39) = -.29$, $p = .07$. Interestingly, however, the Percent of Non-Perseverative Errors did not correlate with *any* of the five WAB variables (all $p > .05$).

Another indicator of WCST performance is the number of trials required to sort the first category. We found no significant correlations between the number of trials for the first sort and any of the WAB variables, all $p > .05$.

Importantly, WAB scores did not correlate with all WCST measures, but rather, the relationships were more specific. Speech and language scores were *not* related to non-perseverative errors or to the number of trials to sort the first category. If there had been a significant correlation with either of these WCST variables, it may have suggested that these patients with aphasia simply did not understand the task. Indeed, all but one patient sorted at least one category, suggesting that a failure to understand the basic task instructions was not a confounding variable. Rather, significant relationships between language measures and conceptual responding and perseveration suggest that patients indeed understood the basic objective of the task, but were impaired in their ability to incorporate feedback to flexibly switch between categories. It is likely that this difficulty also involved an impaired ability to update the sorting strategy.

Also interesting was the fact that fluency did not correlate with any of the WCST measures, suggesting that speech production mechanisms were not related to performance on the WCST. Rather, core language deficits in comprehension and naming were most consistently correlated with impaired WCST performance, along with overall severity.

To determine the specificity of the problem-solving deficit associated with impaired language, we also compared performance on two other tasks, a second problem solving measure (RCPM) and a visuospatial task (Block Design). As with the WCST, performance on the RCPM was also correlated with language performance. Raw scores on the RCPM correlated significantly with AQ, $r(39) = .49$, comprehension, $r(39) = .43$, fluency, $r(39) = .33$, naming, $r(39) = .46$, and repetition, $r\ (39) = .50$, all $p < .05$. To demonstrate discriminant validity, we also compared performance on the five WAB measures with performance on Block Design, a test of visuospatial functioning. None of these correlations even neared significance (all $p > .19$).

There were too few patients to analyze the WCST data based on aphasia classification. Descriptive statistics revealed numerical differences among the subgroups with a considerable amount of variability: mean number of categories sorted (where maximum = 6) in patients with anomic aphasia was 4.2 ($SD = 2.1$), Broca's aphasia 4.0 ($SD = 1.6$), WNL-left hemisphere 3.9 ($SD = 2.1$), WNL-right hemisphere 3.0 ($SD = 2.0$), conduction aphasia 3.2 ($SD = 2.1$), and Wernicke's 0.8 ($SD = 0.5$). The patient who classified as transcortical motor sorted one category, and the patient who was unclassifiable sorted four categories. Not surprisingly, patients with Wernicke's aphasia, who had the most impaired comprehension, had the lowest scores on the WCST (see Fig. 1).

To investigate the brain basis of impaired problem solving in this patient group, lesion data were analyzed using Voxel-based Lesion Symptom Mapping (VLSM; version 1.3, see Bates et al., 2003). VLSM visually represents the relationship between lesion site and performance on a continuous behavioral measure, using tools similar to those in functional neuroimaging. This technique was preferable to separating patients into subgroups (e.g., anterior vs. posterior, right vs. left hemisphere), because of the small sample sizes. Also, VLSM allows one to visualize the relative contributions of various brain regions to a cognitive task, rather than arbitrarily separating patients into "good" and "bad" performers. Voxel-based calculations were performed, examining WCST performance in patients with lesions in each voxel. Fig. 2 shows the resultant VLSM map for overall performance (Percent Correct) on the WCST. "Hotter" areas on the map (i.e., closer to red) represent areas that were more related to impaired performance. The map revealed two major foci, one in the left superior and middle temporal gyri and one in the left posterior inferior frontal gyrus. These regions are represented as the two red areas on the left side of the map. The temporal focus is consistent with the notion that core language abilities (e.g., comprehension) were found to correlate with problem solving performance. The more anterior focus is consistent with an area found to be important for verbal working memory in the functional neuroimaging literature (see Awh et al., 1996). The relationship between working memory and covert verbalization is discussed further in the General discussion.

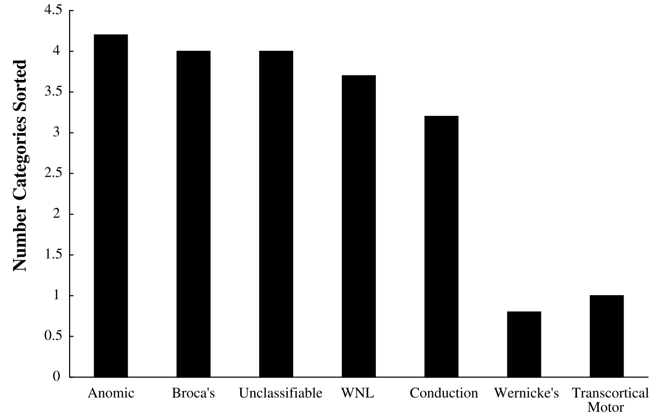

Fig. 1. Number of categories sorted in WCST across aphasia diagnostic subgroups.

3. Experiment 2: Effects of articulatory suppression on WCST performance in normal participants

Experiment 1 suggested that language is involved in performance on tests of non-verbal problem solving. We wanted to further understand the mechanisms involved in this phenomenon by studying WCST performance in normal participants under conditions of articulatory suppression (similar to Hermer-Vazquez et al., 1999; see also Baddeley & Logie, 1999). In Experiment 2, normal undergraduates performed the WCST under one of three conditions: (1) with simultaneous verbal shadowing (saying "na na na"); (2) with simultaneous non-verbal shadowing (rhythmic tapping); or (3) WCST alone. Simultaneous, articulatory suppression has been shown to interfere with on-line verbal processing (e.g., Hermer-Vazquez et al.). If the WCST is dependent on covert language mechanisms, then WCST performance should be significantly reduced in the articulatory suppression condition. Because of the dual task conditions, it was necessary to administer the computer-

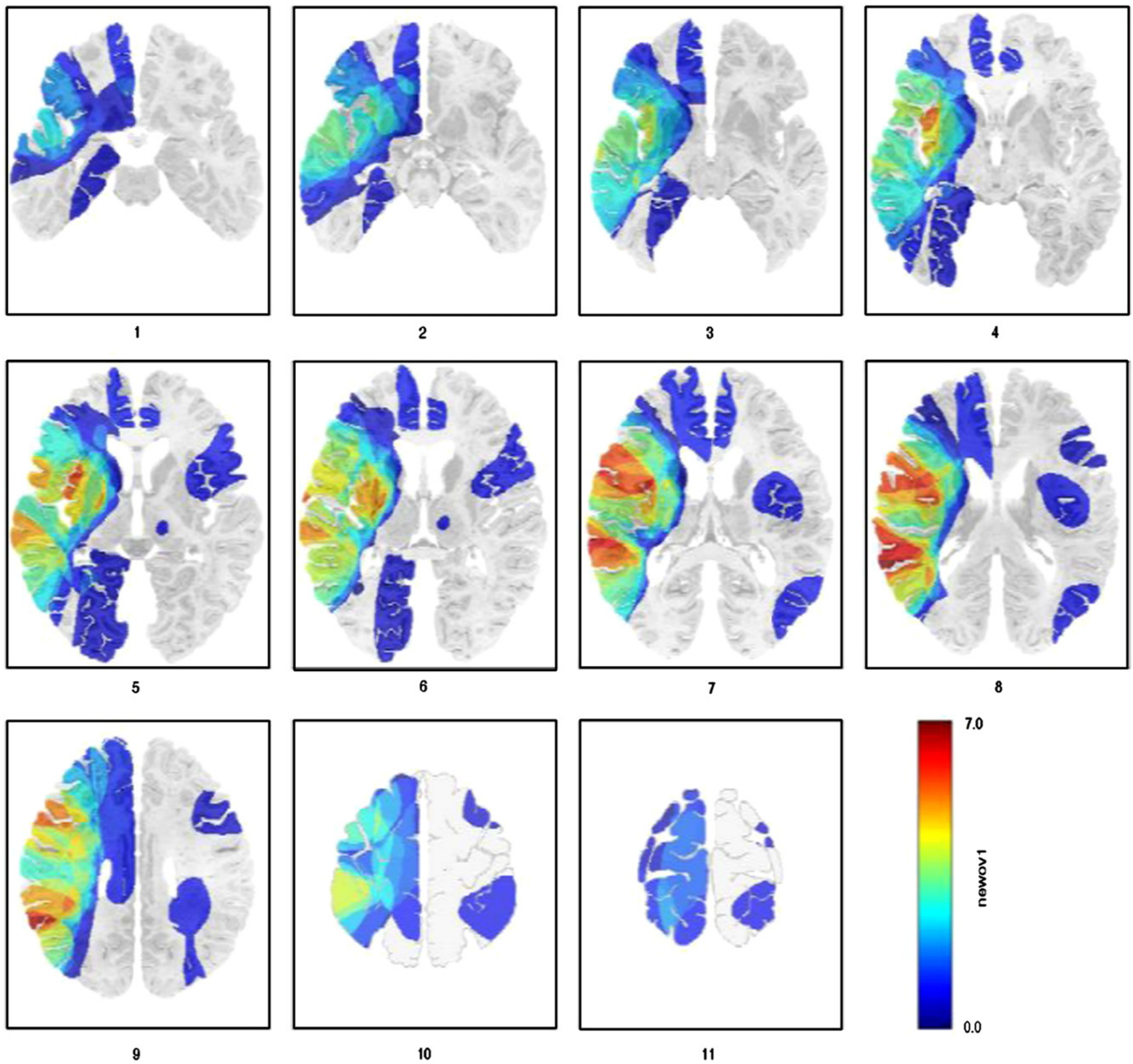

Fig. 2. VLSM map showing the relationship between lesion foci and performance on the WCST (percent correct) across 11 horizontal slices. Patients' raw scores were converted to standard scores (0–1) and then summed at every voxel. "Hotter" areas on the map (i.e., closer to red) represent brain regions that were more related to impaired performance.

ized version of the WCST rather than the standard administration used in Experiment 1. We hoped to model problem-solving deficits observed in aphasic patients in Experiment 1, as well as learn more about the source of those deficits. It was hypothesized that verbal shadowing would lead to impaired WCST performance and more specifically, increased perseveration.

3.1. Method

3.1.1. Participants

Participants were 51 undergraduates (43 women and 8 men) at the Claremont Colleges, recruited primarily from psychology courses. Participants ranged in age from 18 to 22 ($M = 20.0$), and their education level ranged from 12 to 15 years ($M = 13.2$). All participants reported being right-handed except for four left-handed and one ambidextrous individual. An additional six participants were also tested, but were subsequently excluded due to refusal to complete the test ($n = 1$), a speech disorder ($n = 1$), prior familiarity with the WCST ($n = 1$), failure to comply with task instructions ($n = 1$), and performance >2 *SD*s from the group mean ($n = 2$).

All participants read and signed consent forms prior to participation in the study. Participants were offered either extra credit or a candy bar for their participation.

The study was approved by an institutional review board at Scripps College.

3.1.2. Materials

Participants were administered the computerized version of the WCST (Heaton et al., 1993; Psychology Assessment Resources) on a desktop PC. In this version, the four key cards appear across the top of the computer screen, and the choice card appears at the bottom of the screen. The F, G, H, and J keys of the keyboard were labeled with stickers representing the four key cards.

During the WCST procedure, participants heard a rhythm sequence that played on a CD player. The sequence was digitally created with Sibelius (Sibelius Software) and was a sequence of rhythms in 4/4 time that changed every 8 beats and occurred at the rate of 90 beats/min (moderately fast).

3.1.3. Procedure

Participants were tested individually in a noise-attenuated testing room. They were seated in front of a computer and given instructions verbally by the experimenter. Participants were randomly assigned to one of three WCST conditions: verbal shadowing ($n = 17$), non-verbal shadowing ($n = 18$), or baseline ($n = 16$). Participants in the shadowing conditions were first asked to practice shadowing the rhythm sequence for 1 min. Participants in the verbal shadowing condition were asked to say "na" out loud every time there was a beat, and participants in the non-verbal shadowing condition were asked to tap once on the mouse using their non-dominant hand every time they heard a beat. In both shadowing conditions, participants were told not to anticipate the beats but only to shadow them once they were heard. Also, they were told not to worry if they missed a beat or two, but to just keep shadowing as best they could.

All participants were oriented to the monitor for explanation of the WCST procedure. Standardized WCST instructions (Heaton et al., 1993) were used but modified so as to be appropriate for the computerized version. Participants were instructed to use their dominant hand on the keyboard to indicate the key card that they thought the choice card best matched. Feedback was given on the screen in the form of the word "correct" or "incorrect" after each response.

Following the WCST instructions, the rhythmic sequence on the CD was begun. Participants in the shadowing conditions were told that their shadowing responses were being recorded and that it was important for them to perform as accurately as possible on both the shadowing and WCST tasks. Participants in the WCST baseline condition were told that they would hear a rhythmic sequence from the CD player but that they could ignore it.

The examiner remained in the testing room during the procedure. Participants in the shadowing conditions were prompted by the examiner to "keep with the rhythm" if they failed to respond for more than 3 s. The examiner recorded the number of prompts made for each participant. An exclusionary criterion of greater than five prompts was established prior to the study. One participant reached this level and thus was excluded from data analysis.

Participants were debriefed at the end of the study and were given an information sheet to review.

3.2. Results and discussion

WCST data were analyzed with a series of one-way analyses of variance (ANOVA) with Condition (baseline, non-verbal shadowing, or verbal shadowing) as a between-subjects factor. The dependent variables were the same as those analyzed in Experiment 1, namely, Percent Correct, Percent Perseverative and Non-Perseverative Errors, Percent Conceptual Level Responses, and the number of trials to complete the first category. Preliminary analyses revealed no effects of or interactions with gender or handedness, and thus the analyses reported here are collapsed across these variables.

There was a significant effect of Condition on the Percent Correct trials on the WCST, $F(2,48) = 10.60$, $p < .001$. Post hoc Tukey HSD revealed significant differences between verbal shadowing and baseline ($p < .01$) and between non-verbal shadowing and baseline ($p < .05$). There was also a significant effect of Condition on the Percent Perseverative Errors, $F(2,48) = 8.08$, $p < .01$ (see Fig. 3). Post hoc Tukey HSD showed a difference between verbal shadowing and baseline ($p < .01$) and between non-verbal shadowing and base-

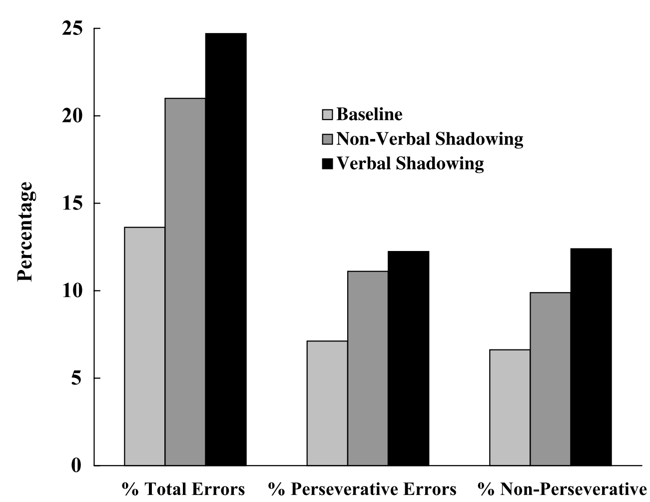

Fig. 3. Performance on the WCST in normal participants in the baseline (WCST alone), non-verbal shadowing (WCST plus tapping), and verbal shadowing (WCST plus articulation) conditions.

line ($p < .05$). The effect of Condition was also significant for Percent Non-Perseverative Errors, $F(2,48) = 6.13$, $p < .01$, and Tukey HSD revealed a difference only between verbal shadowing and baseline ($p < .01$). Last, there was a significant effect of Condition on the Percent Conceptual Level Responses, $F(2,48) = 10.89$, $p < .01$, with significant Tukey HSD for verbal shadowing vs. baseline ($p < .01$) and for non-verbal shadowing vs. baseline ($p < .05$). There was no effect of Condition for the number of trials to complete the first category, $F(2,48) = 2.99$, $p > .05$.

We also ran a MANOVA with Condition as the independent variable and all five of the dependent variables, but the results were identical to the separate one-way ANOVA results described above.

Thus, as predicted, participants in the verbal shadowing condition consistently performed significantly worse than participants in the baseline WCST condition. This finding suggests that articulatory suppression impaired performance. Performance in the verbal shadowing condition, however, was numerically but not statistically worse than that in the non-verbal shadowing condition. It is possible that the use of the rhythmic sequence in the non-verbal shadowing condition may have engaged the auditory-verbal loop to some extent as well. That is, having to listen and make judgments based on the auditory input may have interfered with internal speech or verbal working memory in some participants, even though the output was motoric, namely, tapping. Also, it was not expected that the tendency to make Non-Perseverative Errors would be affected by the Condition manipulation, as Experiment 1 did not show a correlation between this variable and language performance. It is possible that in Experment 2, there was an additional attentional component (in both shadowing conditions) that disrupted performance more generally. This finding and related issues are discussed further below.

4. General discussion

The current study explored the role of language in problem solving. First, we tested neurologic patients with varying degrees of aphasia and found that performance on the Wisconsin Card Sorting Test (WCST) correlated with a number of language measures, most consistently with comprehension and naming scores. Establishing convergent and discriminant validity, we found that another test of problem solving, Raven's Colored Progressive Matrices, also correlated with language performance, but a test of visuospatial skills, the Block Design test, did not. Importantly, the WCST is a nonverbal problem solving task, and thus, the observed relationship suggests that language impairment can manifest as impaired cognition even when overt verbalization is not required. One potential mechanism for this relationship was explored in Experiment 2, in which we tested normal participants on the WCST under conditions of articulatory suppression (Baddeley & Logie, 1999). Normal participants were significantly impaired on the WCST when they were required to vocalize during the task (saying "na na na"), relative to baseline. Together, these findings suggest that covert language processes, more colloquially referred to as "inner speech," may be utilized to support complex problem solving.

In Experiment 1, we found that more severe language impairment was correlated with poor problem solving. However, it was not possible to statistically analyze data across different aphasic classifications (i.e., Broca's, Wernicke's, etc.) due to small group sizes. These classifications are not always useful as they represent syndromes that encompass a constellation of speech and language symptoms. Rather, we focused on specific components of language to determine which were most crucial for problem solving. In particular, we found that comprehension and naming scores correlated most consistently with a number of measures on the WCST, including overall conceptual responding and the degree of perseveration. This finding suggests that core language deficits, such as those seen in severe forms of aphasia (e.g., Wernicke's, global aphasia) are most detrimental to higher cognitive skills such as problem solving. Consistent with this idea, the aphasic subgroup who performed most poorly in our study was patients with Wernicke's aphasia.

Performance in one particular patient with Wernicke's aphasia was very striking. This patient had a doctoral degree and was a well-known scientist and scholar in his field. The WCST no doubt would have been a trivial exercise in logical problem solving for him before his stroke. However, this patient with Wernicke's aphasia was severely impaired on the task. He clearly understood the task instructions, because he responded with an excited "OK" to the examiner's "correct" feedback and with a discouraged "No" to the examiner's "incorrect" feedback. When the target card completely matched the key card, he would say "perfect" *prior* to getting feedback. However, he was completely confounded by the task and performed in the severely impaired range on all WCST measures. Following the task, the patient was asked by the examiner to sort the cards by color, and he did so with ease. Apparently, his severe language deficit prevented him from being able to act upon the examiner's feedback to guide his reasoning on this task.

Many patients with brain injury, especially patients with frontal injury, can sort the first category on the WCST without difficulty, but then have difficulty switching to a new category (Milner, 1963). Such patients will often perseverate, or continue sorting the cards using the same sorting criterion (e.g., color), despite negative feedback that clearly indicates the sorting criterion has changed. Other kinds of errors (e.g., non-perseverative errors) may indicate that the patient is not keeping track of what sorting criterion s/he is using but rather is sorting the cards unsystematically. In the current study, we found that the percentage of perseverative errors (but not non-perseverative errors) was correlated with a number of language measures in the aphasic patients. This finding suggests that flexibility and cognitive switching depend in part on language.

We do not argue here that intact language is necessary for thought in general. Indeed, even aphasic patients with severe language deficits, such as the one described above, execute daily tasks with no apparent confusion, for example, driving, cooking, etc. Rather, we argue that the current findings are consistent with the notion that language representations are involved in certain forms of logical problem solving. As problems become more difficult, they cannot be solved automatically, but require covert verbal mediation or inner speech to buttress on-line problem solving (Sokolov, 1968/1972). The exact nature or locus of this construct—inner speech—has not been clearly defined in the literature. Sokolov has stated that "inner speech emerges as a rather intricate phenomenon, where thought and language are bound in a single, indissoluble complex acting as the speech mechanism of thinking," (p. 1).

A more familiar term that overlaps with the concept of covert verbalization is verbal working memory, which has been suggested to support problem solving along with numerous other on-line cognitive abilities (see Baddeley & Logie, 1999). As Baddeley (2000) observes, people "recode materials verbally so as to take advantage of the capacity of the phonological loop for storing serial order" (p. 420). In the current study, it is possible that the crucial component in both patients with severe aphasia and in normal individuals under conditions of articulatory suppression was a reduced verbal working memory capacity. To test this possibility, we regressed patients' scores on a verbal working memory test (Redfern Repetition Scale; Redfern & Dronkers (unpublished)) with the five different WCST measures. None of these correlations neared significance (all $p > .18$). However, the Repetition subscale on the WAB did correlate significantly with perseveration on the WCST, though not with any of the other WCST measures. Thus, it is possible that verbal working memory plays a role in certain aspects of WCST performance. We would suggest that verbal working memory is a specialized form of covert verbalization that allows one to rehearse/maintain information on-line, whereas covert verbalization is a more general capacity for subvocalizing ongoing thought processes.

As with any correlation, it is possible that some other unmeasured "third" variable could explain the observed relationship between language and problem solving in the current study. That is, some more general working memory or attentional allocation system may underlie both language and problem solving performance. This possibility could potentially explain some of the variability in the WNL patient group in Experiment 1, which included right hemisphere patients, some of whom did not do well on the WCST. It may also explain the observed decrease in performance in normal participants in the non-verbal shadowing condition in Experiment 2.

In the current study, we assessed the role of different brain regions in performance on the WCST using a voxel-based lesion symptom mapping technique (see Bates et al., 2003). We found two lesion foci that were associated with WCST performance, namely, in the left middle and superior temporal gyri and posterior inferior frontal gyri. Some researchers have argued that the presence of cognitive deficits in aphasic patients may not be due directly to the loss of linguistic competence, but rather due to a coincidence of anatomy (Basso et al., 1973, 1981; Hamsher, 1991). Weinstein and Teuber (1957) found that aphasia played a role in cognitive performance on standardized intelligence measures, but that left temporo-parietal patients were still significantly impaired relative to normal controls, even when patients with aphasia were separated out. Thus, both the presence of aphasia and the locus of injury significantly affected performance. In the current study, the correlation between language and problem solving performance would certainly suggest that language decrement is related more directly to effects on other aspects of cognition. Also, the foci detected in the left temporal and inferior frontal gyri are consistent with centers supporting language comprehension and on-line rehearsal of verbal information, respectively (Awh et al., 1996; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004).

Despite the current findings and other literature reviewed, some studies with aphasic patients would seem to challenge the notion that language directly supports reasoning and problem solving (Basso et al., 1973; Kinsbourne & Warrington, 1963; Varley & Segal, 2000). Contrary to the current findings, Basso et al. (1973; see also Basso et al., 1981) reported that performance on the Raven's Matrices test was impaired in both left hemisphere patients with aphasia and right hemisphere patients, and that there was no correlation between Raven's performance and language scores. They did find, however, that aphasics performed poorly on reasoning tests but concluded that this was due to lesions that encroached on other regions critical for cognition. Kinsbourne and Warrington tested two patients with jargon aphasia and reported that they scored in at least the average range on a number of neuropsychological measures. They concluded that "inductive thinking may ... remain essentially intact in spite of gross syntactical disorder" (p. 26). It should be noted, however, that these two patients had mild to no receptive language impairment, which may have contributed to their relatively preserved cognitive abilities. In the current study, comprehension was an important factor in predicting poor problem solving performance. Also, one of the two patients in Kinsbourne and Warrington's study failed a category sorting test, suggesting that problem solving was disrupted to some extent in at least one of the two patients.

More recently, Varley and Segal (2000) reported that an agrammatic patient with a large lesion affecting most of the temporal lobe was able to perform a number of cognitive tasks quite well, including a theory of mind task. They concluded that, while grammar may be crucial to one's ability to attain a certain level of cognitive development, "cognition can operate without grammar" in the adult (p. 726). However, while this patient was poor at grammaticality judgments, he was able to comprehend difficult task instructions in the theory of mind task, suggesting that his language comprehension was only mildly to moderately impaired.

In Experiment 2 of the current study, verbal shadowing or articulatory suppression disrupted performance on the WCST relative to baseline, again supporting the notion that problem solving performance in part is related to subvocalization. However, there was some variability in the verbal shadowing condition, as a few participants appeared much less affected by the articulation suppression. Some of this variability may have been due to individual differences. Interestingly, recent work by Kim (2002) has shown that there may be cross-cultural differences in the use of language to support problem solving. Specifically, Kim found that Asian-American participants were much less likely to use language processes to support problem solving. Kim's work suggests that using verbal mediation to assist problem solving may be in part a product of sociocultural factors. Experiment 2 in the current study included both European- and Asian-American participants, potentially explaining some of the variability in performance that we observed. However, we did not collect ethnic and cultural data and thus were unable to determine the relevance of this variable in the current study.

Another important issue highlighted by the second experiment is that other factors, such as attention, also play a role in problem solving. This was observed as participants in the non-verbal shadowing (tapping) condition also performed worse relative to baseline. The other potential explanation for this effect is that listening and attending to the rhythmic beat tied up auditory-verbal resources to some extent. We would predict that a truly non-verbal condition (e.g., free tapping) would not disrupt problem solving, and current research in our lab is investigating this possibility.

While the current study suggests that problem solving is supported by language, there are a number of cognitive domains that appear unaffected by language impairment. Patients with aphasia demonstrate normal performance on visuospatial tasks such as maze learning (Archibald et al., 1967), tests of personal and extrapersonal space (Semmes, Weinstein, Ghent, & Teuber, 1963), judgment of line orientation (Baldo et al., 2001; Benton, Hannay, & Varney, 1975; Benton, Varney, & Hamsher, 1978), and facial recognition (Baldo et al., 2001; Benton & Van Allen, 1968). These findings are consistent with our failure to find a relationship between Block Design and language performance in aphasic patients in Experiment 1 of the current study.

On a clinical note, it is the case that many aphasic patients are not assessed for cognitive dysfunction outside the domain of language, in part because many of the tools used in neuropsychological assessment require intact speech and language. The failure to fully assess this patient group is problematic, as the current study suggests that language deficits may interfere with other aspects of cognition. Furthermore, these deficits may negatively impact rehabilitation efforts. Some attempts have been made to develop neuropsychological batteries that could be used with language-impaired patients (e.g., the Global Aphasia Neuropsychological Battery; van Mourik, Verschaeve, Boon, Paquier, & van Harskamp, 1992), but such tests are used infrequently.

It has been suggested that the leap in human cognitive capacity relative to other mammals is in part due to the evolution of language (Kuczaj & Hendry, 2003; Sokolov, 1968/1972). Language being a symbolic representation system allows us to not simply represent concepts, but more importantly for problem solving, facilitates our ability to manipulate those concepts and generate novel solutions. Moreover, the ability to internalize language in the form of inner speech allows us to manipulate concepts and solve problems covertly. As shown in the current study, when language is disrupted, either by neurologic insult or verbal interference, we are left with a less sophisticated, less flexible capacity for analyzing and solving complex problems.

References

Archibald, Y. M., Wepman, J. M., & Jones, L. V. (1967). Nonverbal cognitive performance in aphasic and nonaphasic brain-damaged patients. *Cortex, 3*, 275–294.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koepee, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science, 7*, 25–31.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences*, 4*, 417–423.

Baddeley, A., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Baldo, J. V., Delis, D., Wilkins, D., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Archives of Clinical Neuropsychology, 19*, 407–419.

Baldo, J. V., Elder, J., Larsen, J., Dronkers, N. F., Redfern, B., & Ludy, C. (2001). Is cognition intact in patients with aphasia?. *Brain and Language, 79*, 64–67 (Abstract).

Basso, A., Capitani, E., Luzzatti, C., & Spinnler, H. (1981). Intelligence and left hemisphere disease: the role of aphasia, apraxia, and size of lesion. *Brain, 104*, 721–734.

Basso, A., De Renzi, E., Faglioni, P., Scotti, G., & Spinnler, H. (1973). Neuropsychological evidence for the existence of cerebral areas critical to the performance of intelligence tasks. *Brain, 96*, 715–728.

Bates, E., Wilson, S., Saygin, A. P., Dick, F., Sereno, M., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6*, 448–450.

Benton, A., Hannay, H. J., & Varney, N. R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology, 25*, 907–910.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Cortex, 4*, 344–358.

Benton, A. L., Varney, N. R., & Hamsher, K. D. (1978). Visuospatial judgement: A clinical test. *Archives of Neurology, 35*, 364–367.

Borod, J., Carper, M., & Goodglass, H. (1982). WAIS performance IQ in aphasia as a function of auditory comprehension and constructional apraxia. *Cortex, 18*, 199–210.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*, 6.

De Renzi, E., Faglioni, P., Savoiardo, M., & Vignolo, L. (1966). The influence of aphasia and of the hemispheric side of the cerebral lesion on abstract thinking. *Cortex, 2*, 399–420.

Dronkers, N., Wilkins, D., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*, 145–177.

Edwards, S., Ellams, J., & Thompson, J. (1976). Language and intelligence in dysphasia: Are they related?. *British Journal of Disorders of Communication, 11*, 83–94.

Gentner, D. (2003). Why we're so smart. In D. Genter & S. Goldin-Meadow (Eds.), *Language in Mind*. Cambridge: MIT Press.

Hamsher, K. (1991). Intelligence and aphasia. In M. Sarno (Ed.), *Acquired aphasia* (2nd ed.). San Diego: Academic Press.

Heaton, R., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin card sorting test (WCST) manual revised and expanded*. Odessa, FL: Psychological Assessment Resources.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual task studies of space and language. *Cognitive Psychology, 39*, 3–36.

Hjelmquist, E. K. (1989). Concept formation in non-verbal categorization tasks in brain-damaged patients with and without aphasia. *Scandiniavian Journal of Psychology, 30*, 243–254.

Hurlburt, R. T. (1990). *Sampling normal and schizophrenic inner experience*. New York: Plenum Press.

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103*, 548–560.

Kertesz, A. (1982). *Western aphasia battery*. New York: Grune & Stratton.

Kertesz, A., & McCabe, P. (1975). Intelligence and aphasia: Performance of aphasics on Raven's coloured progressive matrices (RCPM). *Brain and Language, 2*, 387–395.

Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *Journal of Personality and Social Psychology, 83*, 828–842.

Kinsbourne, M., & Warrington, E. K. (1963). Jargon aphasia. *Neuropsychologia, 1*, 27–37.

Klouda, G. V., & Cooper, W. E. (1990). Information search following damage to the frontal lobes. *Psychological Reports, 67*, 411–416.

Kuczaj, S. A., & Hendry, J. L. (2003). Does language help animals think. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind*. Cambridge: MIT Press.

Larrabee, G., & Haley, J. (1986). Another look at VIQ-PIQ scores and unilateral brain damage. *International Journal of Neuroscience, 29*, 141–148.

Miller, L. A., & Tippett, L. J. (1996). Effects of focal brain lesions on visual problem-solving. *Neuropsychologia, 34*, 387–398.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9*, 100–110.

van Mourik, M., Verschaeve, M., Boon, P., Paquier, P., & van Harskamp, F. (1992). Cognition in global aphasia: Indicators for therapy. *Aphasiology, 6*, 491–499.

Piercy, M. (1964). The effects of cerebral lesions on intellectual function: A review of current research trends. *British Journal of Psychiatry, 110*, 310–352.

Raven, J. (1962). *Coloured progressive matrices*. New York: The Psychological Corporation.

Redfern, B., & Dronkers, N. F. (unpublished). *Redfern Repetition Rating Scale*.

Semmes, J., Weinstein, S., Ghent & Teuber, H. (1963). Correlates of impaired orientation in personal and extrapersonal space. *Brain, 86*, 747–772.

Sokolov, A. N. (1968/1972). *Inner speech and thought* (G.T. Onischenko, Trans.) New York: Plenum Press, Original work published 1968.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63*, 289–298.

Varley, R., & Segal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology, 10*, 723–726.

Weinstein, S., & Teuber, H. (1957). Effects of penetrating brain injury on intelligence test scores. *Science, 125*, 1036–1037.