Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury?

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Journal of Clinical and Experimental Neuropsychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td></td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Kave, Gitit; Open University, Israel, Department of Education and Psychology</td>
</tr>
<tr>
<td>Keywords:</td>
<td>verbal fluency, traumatic brain injury, switching and clustering, executive functions, language testing</td>
</tr>
</tbody>
</table>
Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury?

Gitit Kavé¹ AND Eyel Heled²,³

Eli Vakil⁴

Eugenia Agranov³,⁵

¹Department of Education and Psychology, The Open University, Raanana, Israel;
²Department of Psychology, Hebrew University, Jerusalem, Israel; ³Sheba Medical Center, Ramat Gan, Israel; ⁴Department of Psychology and Leslie and Susan Gonda (Goldschmied) Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel; ⁵Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Address correspondence to:

Gitit Kavé

Department of Education and Psychology
The Open University
1 University Road, P.O. Box 808, Ra'anana 43107, Israel
Tel: 972-9-778-1445, Fax: 972-9-778-0632
Email: gkave@012.net.il
Abstract

This paper examines switching and clustering in phonemic and semantic fluency tasks in individuals with traumatic brain injury (TBI). Fluency tasks were administered to 30 Hebrew-speaking patients with TBI and 30 age-matched control participants. Significant group differences were found in total output, number of switches, and number of clusters on both tasks, but not in mean cluster size. Unlike prediction, \( z \)-scores of the number of semantic switches and clusters were lower than the equivalent \( z \)-scores on the phonemic test. Results highlight the executive component of semantic fluency and the importance of using this task when assessing cognitive functioning after TBI.

**Key words:** verbal fluency, traumatic brain injury, switching and clustering, executive functions, language testing, Hebrew
Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury?

Tests of verbal fluency are commonly used in neuropsychological assessment of individuals with brain damage (Henry, Crawford, & Phillips, 2004), contributing to the evaluation of executive functioning. These tests ask individuals to generate as many different words as possible that follow a certain rule in a limited time frame, measuring the ability to search lexical stores, to retrieve information from semantic memory, as well as to switch cognitive set (Henry & Crawford, 2004a, 2004b; Kavé, 2005; Kavé, Avraham, Kukulansky-Segal, & Herzberg, 2007; Kavé, Kigel, & Kochva, 2005; Troyer, 2000). It has been suggested that phonemic fluency (i.e., retrieving words that begin with a certain letter) might be more impaired in individuals with frontal brain damage (Rosser & Hodges, 1994; Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998), whereas semantic fluency (i.e., retrieving words that belong to a given semantic category) might be more impaired in individuals with temporal brain damage. Moreover, some authors believe that the phonemic task should be more sensitive to traumatic brain injury (TBI) than the semantic fluency task (Capitani et al., 2009; Juardo, Mataro, Verger, Bartumeus, & Junque, 2000). However, not all studies support this conclusion (Henry & Crawford, 2004a, 2004b). The current research looks at a quantitative and qualitative analysis of fluency variables, aiming to elucidate the cognitive components most sensitive to TBI.

Previous research on individuals with brain injury has demonstrated that their word production on fluency tests is limited (Axelrod, Tomer, Fisher, & Aharon-Peretz, 2001; Bittner, & Crowe, 2006, 2007; Busch, McBride, Curtiss, & Vanderploeg, 2005; Capitani et al., 2009; Henry & Crawford, 2004a; Juardo et al., 2000). A meta-analysis of studies of TBI patients shows that the magnitude of the deficit cannot be accounted for by a failure to match
patients and control participants on pre-morbid IQ, current VIQ, education, or psychomotor speed (Henry & Crawford, 2004a). In addition, impairments in word finding are not the principal cause of the reduction in fluency output, as patients with and without naming difficulties demonstrate equally deficient fluency (Bittner & Crowe, 2006). Instead, it is assumed that patients with TBI are impaired on fluency tests due to executive dysfunction, most notably due to difficulties in self-generative behavior (Busch et al., 2005). According to Capitani et al. (2009) as well as Juardo et al. (2000), TBI leads to greater deficits in phonemic rather than semantic fluency, probably due to its impact on the frontal lobes. However, TBI most often leads to diffuse damage in frontal as well as temporal brain regions (Vakil, 2005), and hence both fluency tasks might be impaired.

Word retrieval on all verbal fluency tests depends on lexical knowledge as well as on effective search processes that require set shifting (Troyer, Moscovitch, & Winocur, 1997). A meta-analysis of studies of persons with Alzheimer's disease, who have severe lexical-conceptual disorders, demonstrated greater difficulties in semantic fluency relative to phonemic fluency (Henry et al., 2004). Furthermore, focal temporal damage has been associated with a lesser deficit on phonemic fluency and a larger deficit on semantic fluency (Henry & Crawford, 2004b). Although persons with temporal lesions show a semantic fluency deficit, this deficit is often found in individuals with frontal lesions as well (Rogers, Ivanoiu, Patterson, & Hodges, 2006). It has been shown that persons with focal frontal lobe lesions produce significantly fewer words on phonemic fluency tests than do healthy controls, and also perform worse on that test than do persons with non-frontal lesions (Alvarez & Emory, 2006). Yet, according to Henry and Crawford (2004a, 2004b), frontal brain damage, whether focal or diffuse, leads to a comparable impairment on both phonemic and semantic tests.
Some authors have used qualitative methods to examine the cognitive strategies underlying verbal fluency tasks (e.g., Fagundo, López, Romero, Guarch, Marcos, & Salamero, 2008; Kavé et al., 2007, 2008; Koren, Kofman, & Berger, 2005; Kosmidis, Vlahou, Panagiotaki, & Kiosseoglou, 2004; Lanting, Haugrud, & Crossley, 2009; Rosselli, Tappen, Williams, Salvatierra, & Zoller, 2009; Troyer et al., 1997; Woods et al., 2004). Instead of comparing the total number of words provided on the phonemic and semantic tasks, these analyses have focused on two components termed *switching* and *clustering*. According to Troyer et al. (1998), when generating words on the phonemic and semantic fluency tasks, participants produce clusters of phonemically or semantically related words and once a subcategory is exhausted they switch to another subcategory. Thus, performance on these tasks relies on (1) an executive component (i.e., switching) responsible for strategic search, response initiation, monitoring, shifting, and flexibility; and (2) an associate component (i.e., clustering) that reflects the semantic organization of memory stores (Troyer et al., 1997; Troyer et al., 1998; Troyer, 2000).

It is further assumed that anterior brain regions play a more important role in switching than in clustering. Troyer et al. (1998) examined this hypothesis in persons with focal brain lesions, finding individuals with frontal lobe lesions to switch less frequently than healthy participants and to produce normal cluster size on both the phonemic and the semantic tasks. In contrast, individuals with temporal lobe lesions exhibited normal switching and clustering on the phonemic task, but were impaired in switching on the semantic task. Although persons with temporal lobe lesions showed no marked deficit in cluster size, those who had left temporal lesions produced smaller clusters than those who had right temporal lesions. This study suggested that phonemic clustering was less dependent on the integrity of lexical stores than was semantic clustering and that the most discriminating index among the patient groups...
was the number of switches on the phonemic fluency task, which was impaired only in persons with frontal lesions.

Additional evidence supporting the assumption that switching is an executive function, whereas clustering is more dependent on lexical abilities, especially within the semantic task, comes from research of various non-focal neuropsychological disorders, such as dementia due to HIV (Woods et al., 2004), adults with ADHD (Tucha et al., 2005), multiple sclerosis (Troster et al., 1998), as well as depression (Fossati, Guillaume, Ergis, & Allilaire, 2003). On the other hand, individuals with Alzheimer's disease produce smaller clusters than normal (Epker, Lacritz, & Cullum, 1999; Fagundo et al., 2008; Troster et al., 1998). In addition, persons with schizophrenia, who suffer from disproportionate semantic fluency impairment relative to phonemic fluency (Bokat & Goldberg, 2003; Kremen, Seidman, Faraone, & Tsuang, 2003), show a disproportionate decrease in the number of clustered words (Bozikas, Kosmidis, & Karavatos, 2005).

Thus, the examination of switching and clustering can clarify the relative contribution of executive strategies and semantic stores to fluency performance in TBI. Since TBI primarily affects frontal regions, we predict that measures of executive functioning (e.g., the number of switches as well as the number of clusters) will be more impaired in this population than measures that represent semantic storage (e.g., cluster size). We want to examine whether executive deficits will be comparable on the phonemic and semantic tasks or more pronounced on one task relative to the other. The aim of the current paper, then, is to determine whether qualitative measures of verbal fluency contribute to our understanding of the underlying mechanisms that lead to fluency deficits in TBI.

Method
Participants

Sixty native Hebrew speakers participated in the study, 30 in the TBI group and 30 in the control group, each group consisting of 22 men and 8 women. The TBI group was recruited at the day treatment brain injury unit of the Sheba medical center, Tel Hashomer, Israel. Patients had moderate to severe TBI, with inclusion criteria consisting of 6-24 months post injury, a Glasgow Coma Scale (GCS) score of 3-12, loss of consciousness of at least 30 minutes, age between 18-35, and at least 12 years of education. Exclusion criteria included previous brain damage, psychiatric disorder prior to injury, combination of TBI and spinal cord injury, as well as aphasia. Radiological findings showed frontal and temporal contusions, epidural and subdural hematomas, and diffuse axonal injury. Table 1 summarizes background information for patient and control groups. Post-traumatic amnesia (PTA) was determined for every patient, operationally defined as the time when the patient was oriented to self, place and time, remembered information presented by the examiner, and recognized three objects (as in Shores, Marosszeky, Sandanam, & Batchelor, 1986). All questions addressed at the patient had to be answered correctly for a three-day period and PTA was counted up to the first of these three days. Verbal IQ was available for 22 (73%) of the patients, and was within normal limits in all cases.

The control group was selected from Kavé et al.’s (2007) sample to match the TBI group in age and gender. Healthy participants were recruited through places of employment, university classes, and word of mouth. Persons with a known history of learning disorders, psychiatric disturbances, neurological disease, or head trauma were not included in the study. No verbal IQ scores were available for the healthy participants. The two groups did not differ...
in age, $t(58) = -.164, \text{ ns}$, but the TBI group was slightly less educated than the control group, $t(58) = -2.298, p < .05$ (see Table 1), as expected by their age of injury. Participant recruitment in both groups was conducted in accordance with institutional research guidelines.

**Procedure**

Participants were asked to provide as many words as possible within 60 seconds on each of three letters (phonemic test) and three categories (semantic test). The phonemic fluency test was administered first and then the semantic fluency test and the order of letters, as well as the order of semantic categories, was constant across participants. Responses were written verbatim, with errors or repetitions subsequently excluded from the total score. When a questionable response was provided, clarifications were invited at the end of the one-minute interval.

**Phonemic fluency** was assessed by obtaining the number of words generated in one minute for the letters *bet* (/b/), *gimel* (/g/), and *shin* (/š/). Instructions were as follows: "I want you to say as many Hebrew words as possible that begin with a certain letter. You may say any word except for names of people and places, such as *Tomer* or *Tel Aviv*. Also, you should use different words rather than the same word with a different ending. For example, if you say *tapuz* ('orange'), don't also say *tapuzim* ('oranges'). If you say a verb, use the simplest form *halax* ('he went') and not *halaxti* ('I went') or *holex* ('he goes'). Please don't say words that are attached to other words, such as *mi-shamayim* ('from the sky') or *la-kise* ('to the chair')."

**Semantic fluency** was assessed by obtaining the number of words generated in one minute for each of the following three semantic categories: animals, fruits and vegetables, and vehicles. Fruits and vegetables were treated as one category in order to avoid the ambiguity between botanical definitions and common usage (as in 'avocado'). It was specified that for
the category of vehicles only types of transportation should be provided while brand names were unacceptable.

Scoring

When homophones were provided, the second mention was counted only if the participant pointed out the alternate meaning explicitly (i.e., gamal 'camel', 'repaid'). Words inflected in both masculine and feminine forms (e.g., gever-gveret 'mister-mistress'; sus-susa 'horse-mare') were counted as one, whereas an animal and its offspring were counted as separate words (e.g., para 'cow' and egel 'calf'). Synonyms were counted as two (matos and aviron 'airplane'). Names of subcategories on the semantic test (e.g., bird) were not given credit if specific items within that subcategory (e.g., dove, eagle) were also provided. Slang terms (e.g., shluk 'sip'), as well as foreign words (e.g., bandana, gangster), were generally acceptable.

Clustering and switching: In line with Troyer et al.’s (1997) guidelines, repetitions and mistakes were included in the scoring of clustering and switching. An item that appeared in two clusters was coded in both. For example, the "cat" in "dog, cat, tiger, lion" was counted both as part of the cluster of pets and as part of the cluster of felines. In cases in which a small cluster was embedded within a larger cluster, only the larger cluster was counted. Thus, if farm birds were generated among other farm animals, only one cluster was counted (i.e., horse, cow, chicken, duck, turkey = 1 cluster). Semantic clusters generated within the phonemic task, as well as phonemic clusters generated within the semantic task, were not scored.

Phonemic clustering: a cluster was counted when two consecutive words shared the first consonant and vowel (gezer-geshem), shared the first and second consonant but differed
in the vowel of the opening syllable (\textit{gina-ganav}), rhymed (\textit{shamayim-shinayim}), or included duplication (\textit{barbur-bilbul}).

**Semantic clustering:** where possible, subcategories were based on previous studies (Troyer et al., 1997; Kosmidis et al., 2004). Guidelines were formulated for consistency sake but flexibility was allowed for the coding of associated words that did not fall under the list of pre-defined clusters. On the \textit{Animal} category, clusters were coded according to habitat, zoological family, and family relation, which were further classified into relevant subcategories. On the \textit{Fruit and Vegetable} category, clusters were coded according to either fruits or vegetables, with subcategories further defined by season, botanical family, manner of eating, etc. On the \textit{Vehicle} category, clusters were coded according to land, water, or air-borne means of transportations, with further classifications within land vehicles defined by common use (see Kavé et al., 2008, for further details of subcategories).

Four variables were derived for each fluency test on the basis of the aforementioned criteria:

1. **Total fluency score:** all words excluding repetitions and errors, summed across the three letters for the phonemic task and across the three categories for the semantic task.

2. **Mean cluster size:** Following Troyer et al. (1997), the number of words in a cluster was counted from the second word. That is, a cluster of two words was coded as 1; a cluster of three words was coded as 2, and so forth. A mean of all clusters of two words or more was computed for every person for each letter or semantic category. These means were then averaged across the three letters to yield the mean phonemic cluster size of each participant, and across the three semantic categories to yield the mean semantic cluster size of each participant.

3. **Number of switches:** The number of switches between clusters of two words or more, between a cluster and a single word generated outside a cluster, and among those out-
of-cluster single words (as in Troyer et al., 1997), was counted for every person for each letter and for each semantic category. Switches produced by each participant were summed across the three letters to yield the total phonemic number of switches score, and across the three semantic categories to yield the total semantic number of switches score.

4. Number of clusters: The number of clusters was counted separately, without single words, in order to examine participants’ use of word association. As noted by Koren et al. (2005), the presence of single words may indicate that participants are in fact unable to utilize an associative strategy and thus a measure that leaves out the single words is essential when focusing on the tendency to produce related words.

One of the raters who coded the data in Kavé et al. (2007) also coded all patient data. In order to calculate inter-rater reliability, another rater coded the responses generated by 10 participants with TBI. On the phonemic task, correlations between the two raters were $r = .91$, $p < .01$, for mean cluster size, $r = .97$, $p < .01$, for number of switches, and $r = .84$, $p < .01$, for number of clusters. On the semantic task, correlations were $r = .97$, $p < .01$, for mean cluster size, $r = .96$, $p < .01$, for number of switches, and $r = .81$, $p < .01$, for number of clusters.

Control data were not coded again but rather taken from Kavé et al. (2007). Inter-rater reliability for control data was based on correlations between two judges who coded 30 protocols out of 100. All correlations between these two judges were high and significant at the .01 level (see Kavé et al., 2007, for exact details).

Results

To rule out the possibility that the higher education level of the control group accounted for the current results, correlation analyses were conducted between years of education and fluency scores. Within the control group alone, only phonological cluster size was significantly related to education, $r = -.403$, $p < .05$, with larger clusters in less educated...
individuals. The other seven correlations were non-significant ($p > .05$). Within the TBI group, none of the eight measures was significantly correlated to education. When correlations were examined for the two groups together, the number of switches on the phonological task was significantly correlated with education, $r = .378, p < .05$, with more switches produced by more highly educated individuals. No other correlations were statistically significant, and therefore we decided not to enter years of education as a covariate in the analyses described below.

Table 2 displays means and standard deviations of fluency performance for each group. Because each variable was scored on a different scale, we chose not to run a factorial mixed design with repeated-measures comparisons that would look at all variables together. Instead, raw scores were analyzed by eight independent-samples t-tests, conducted for each variable separately. A Bonferroni correction was used to adjust the level of significance, and each comparison was examined against a significance level of .00625 (.05/8). The TBI group produced significantly fewer words than did the control group on both the phonemic task, $t(58) = -7.917, p < .001$, Cohen's $d' = -2.04$, and the semantic task, $t(58) = -8.490, p < .001$, Cohen's $d' = -2.19$. No significant group differences were found for mean phonemic cluster size, $t(58) = -2.083, p = .042$, Cohen's $d' = -.53$, or for mean semantic cluster size, $t(58) = -2.637, p = .011$, Cohen's $d' = -.67$. Individuals with TBI produced significantly fewer phonemic switches than did healthy participants, $t(58) = -7.522, p < .001$, Cohen's $d' = -1.88$, as well as significantly fewer semantic switches, $t(58) = -9.161, p < .001$, Cohen's $d' = -2.37$. The TBI group also produced significantly fewer clusters than did the control group on both the phonemic task, $t(58) = -5.897, p < .001$, Cohen's $d' = -1.52$, and the semantic task, $t(58) = -6.929, p < .001$, Cohen's $d' = -1.79$. Note that the strongest effect size was found for the comparison of the number of semantic switches.
To determine the sensitivity and specificity of both tasks, we compared scores in both groups to age-appropriate normative scores that did not include the healthy participants of the current study (Kavé, 2005), looking at total scores only as there are no normative data for the qualitative variables. We chose a cutoff score of 1.5 standard deviations below the mean normative score and counted how many individuals in each group scored outside that range. This analysis showed that 19 patients scored below the 1.5 SD cutoff score on the phonemic test, setting the sensitivity of the test at 63%, whereas the semantic test showed sensitivity of 80% (i.e., 24 patients scored below the cutoff). Specificity was high for both tests: 96% for the phonemic test (that is, 29 out of 30 healthy individuals scored above the cutoff), and 93% for the semantic test (that is, 28 out of 30 healthy individuals scored above the cutoff).

Next we computed z-scores for each participant on all fluency variables using the mean and standard deviation of the control group in the current study. Table 3 presents the means of z-scores for each of the eight variables. Setting a cutoff score of 1.5 standard deviations below the mean of the control group identified 26 patients (87%) as impaired on the total semantic score, and 21 patients (70%) as impaired on the total phonemic score (see Table 3). Similar asymmetries between the semantic and phonemic tasks were found for the number of switches (83% vs. 70% patients below cutoff, respectively), as well as the number of clusters (60% vs. 40% patients below cutoff, respectively). Mean z-scores of cluster size in the TBI group indicated no impairment on either fluency test.
After standardizing all scores within the TBI group according to the mean score of the control group in the current study, we plotted $z$-scores by variable in order to compare impairments across fluency tasks using the same scale. Figure 1 shows that the mean standardized score of the number of semantic switches was the lowest of all $z$-scores. In general, standardized scores on the semantic test were lower for all measures than standardized scores on the phonemic test, except for mean cluster size, which was equal across both tasks.

Finally, to determine whether the phonemic or the semantic test best discriminated between groups, we conducted four repeated-measures analyses of variances on standardized scores, with task (phonemic, semantic) as the within-subject variable and group (TBI, control) as the between-subject variable. As variables are inherently dependent on each other, so that by definition the number of switches or clusters is defined by the number of words, we chose to conduct a separate analysis for each variable. An analysis that compared standardized total scores found a significant main effect of group, $F(1, 58) = 81.949, p < .001, \text{Eta}^2 = .586$, but no significant main effect of task, $F(1, 58) = 3.200, \text{ns}$, and no significant interaction $F(1, 58) = 3.212, \text{ns}$. An analysis that compared standardized scores of number of switches found a significant main effect of group, $F(1, 58) = 99.938, p < .001, \text{Eta}^2 = .633$, a significant main effect of task, $F(1, 58) = 6.742, p < .05, \text{Eta}^2 = .104$, as well as a significant interaction $F(1, 58) = 6.812, p < .05, \text{Eta}^2 = .105$. As can be seen in Table 3, the difference between mean $z$-score of phonemic switches and mean $z$-scores of semantic switches was greater in the TBI group than it was in the control group, showing a greater deficit in switching on the semantic task. An analysis that compared standardized scores of the number of clusters found a
significant main effect of group, \( F(1, 58) = 61.121, p < .001, \eta^2 = .513 \), a significant main effect of task, \( F(1, 58) = 4.369, p < .05, \eta^2 = .070 \), as well as a significant interaction \( F(1, 58) = 4.349, p < .05, \eta^2 = .070 \). These results corroborate the results of the analysis of the number of switches, showing a greater deficit on the semantic relative to the phonemic task within the TBI group. An analysis that compared standardized scores of cluster size found a significant main effect of group, \( F(1, 58) = 8.218, p < .05, \eta^2 = .124 \), but no significant main effect of task, \( F(1, 58) = .004, ns \), as well as no significant interaction \( F(1, 58) = .002, ns \). That is, patients made significantly smaller clusters than did control participants across tasks, but there was no difference between the size of clusters on the phonemic and the semantic tests.

**Discussion**

Contrary to prediction, individuals with TBI demonstrated a greater deficit on the semantic fluency task than on the phonemic fluency task, yet their deficit reflects difficulties in switching rather than in clustering, thus attesting to an underlying executive dysfunction. Previous studies suggested that phonemic fluency might be more impaired in individuals with TBI than semantic fluency (Capitani et al., 2009; Juardo et al., 2000; Rosser & Hodges, 1994; Troyer et al., 1998), because the phonemic task presumably relies more heavily on frontal brain structures, whereas the semantic task relies more heavily on temporal brain regions. Nevertheless, this suggestion has not received consistent support in the literature, perhaps because TBI is not necessarily confined to frontal regions (Vakil, 2005). In fact, in their meta-analyses of earlier studies, Henry and Crawford (2004a, 2004b) showed that frontal brain damage leads to comparable impairment on both phonemic and semantic tests. Furthermore, Kavé et al. (2007) found that scores on both fluency tasks were equally associated with performance on a test of mental flexibility (the Homophone Meaning Generation Test, HMGHT) through a common executive component.
Both fluency tasks involve strategic search, response initiation, monitoring, shifting, and flexibility, which can be measured through the number of switches (Troyer et al., 1997), as well as the number of clusters (Kavé et al., 2008; Koren et al., 2005), whereas the size of the clusters generated by participants reflects reliance on lexical stores. Indeed, patients with TBI are known to be impaired on a variety of executive tasks that include self-generative behavior, flexibility and set shifting, mental control, and self monitoring (Busch et al., 2005). It is not surprising, then, that our patients were found to be impaired on both fluency tasks, as well as on both variables that measured set shifting (i.e., number of switches and number of clusters), but no deficit in cluster size was seen in this group. Note, though, that the current results did not reveal entirely comparable deficit on both fluency tests. Instead we found that patients were more impaired on the semantic test than on the phonemic test.

Our findings suggest that the semantic test is more sensitive to TBI than is the phonemic test, when scores are compared either to normative data or to data of matched control participants. Moreover, the number of switches and the number of clusters were more affected by brain injury on the semantic task relative to the phonemic task. In fact, the mean standardized score of the number of clusters on the phonemic task was within normal range. Importantly, the variable that led to the lowest standardized scores was the number of switches on the semantic task. It appears, then, that the semantic test is sensitive to the difficulties caused by TBI specifically because it taps into executive processes.

Note that phonemic fluency is most typically assessed through a sum score of responses to three letters (FAS in English; b, g, sh in Hebrew), yet this is not always the case for semantic fluency, which has been commonly evaluated with the single category of 'animals' (e.g., Curtis, Thompson, Greve, & Bianchini, 2008; Epker et al., 1999; Fagundo et al., 2008; Fossati et al., 2003, among many others). It is possible that the current results reflect the fact that we used three semantic categories rather than just one. Indeed, when the category of
animals alone was examined in the current study, the effect size of the group difference on this category was not only smaller than the effect found for the three semantic categories together (Cohen's $d' = -1.93$ for animals alone vs. Cohen's $d' = -2.19$ for all three categories), but also slightly smaller than the effect found for the three letters together (Cohen's $d' = -1.93$ for animals alone vs. Cohen's $d' = -2.04$ for all three letters). Note that test-retest reliability of the three-letter version is better than reliability for any one letter (Harrison, Buxton, Husain, & Wise, 2000), most likely because sampling a greater range of behavior improves the validity of results. It is safe to assume that sampling word output of three categories improves the validity of the semantic test relative to sampling one category alone, especially since the inter-individual variance is large within both healthy and brain-injured groups of participants. Thus, it could be the case that previous findings of a greater deficit in phonemic relative to semantic fluency reflected in part the fact that the former was assessed over three letters (and three minutes), whereas the latter was assessed with only one category.

One major limitation of our study design must be acknowledged. Had we used other cognitive tests we could have better characterized the underlying cognitive processes leading to the observed impairment. Unfortunately, we administered no other executive tests to all participants and had no independent measures of processing speed. While these data were available for some patients, the information was not available for all patients or for the healthy participants. Thus, we cannot rule out the possibility that the differences recorded here reflect compromised speed of processing rather than a deficit in executive abilities. Note however, that a decrease in processing speed would most probably have had a comparable impact on both fluency tasks. Notwithstanding these limitations, our results point to the need to assess semantic fluency in individuals with brain damage as a measure of executive functioning rather than to use phonemic fluency alone. We also believe that our findings demonstrate the importance of assessing semantic fluency with more than just one category.
Acknowledgements

This study was partially supported by a grant from the Rehabilitation Center for Veterans after traumatic brain injury (TBI) in Jaffa, Israel, given to the third author. We wish to thank Ayelet Avraham and Reut Lerer for their help in coding the data. There are no financial or other relationships that could be interpreted as a conflict of interests.
Verbal fluency in TBI

References


Table 1

Sample characteristics by group

<table>
<thead>
<tr>
<th></th>
<th>TBI group (N = 30)</th>
<th>Control group (N = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>19-35</td>
<td>25.47 (4.69)</td>
</tr>
<tr>
<td></td>
<td>19-35</td>
<td>25.67 (4.74)</td>
</tr>
<tr>
<td><strong>Education (years)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>12-16</td>
<td>12.63 (1.33)</td>
</tr>
<tr>
<td></td>
<td>12-16</td>
<td>13.47 (1.48)</td>
</tr>
<tr>
<td><strong>Time post injury (months)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>12.47 (5.39)</td>
</tr>
<tr>
<td><strong>GCS</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>3-12</td>
<td>6.00 (2.44)</td>
</tr>
<tr>
<td><strong>LOC (days)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>0.5-90</td>
<td>17.45 (19.01)</td>
</tr>
<tr>
<td><strong>PTA (days)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>4-180</td>
<td>51.63 (44.55)</td>
</tr>
<tr>
<td><strong>VIQ (N = 22)</strong></td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>82-120</td>
<td>102 (11.62)</td>
</tr>
</tbody>
</table>

Note: GCS = Glasgow coma scale; LOC = loss of consciousness; PTA = post-traumatic amnesia; VIQ = WAIS-III verbal IQ.
Table 2

Means, standard deviations and range of scores of the eight fluency variables, by group

<table>
<thead>
<tr>
<th>Fluency test</th>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TBI group (N = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control group (N = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic</td>
<td>Total 3 letters</td>
<td>25.80</td>
<td>8.21</td>
<td>9-40</td>
<td>45.50</td>
<td>10.87</td>
<td>27-68</td>
</tr>
<tr>
<td></td>
<td>Mean cluster size</td>
<td>1.31</td>
<td>.41</td>
<td>.67-2.67</td>
<td>1.51</td>
<td>.30</td>
<td>1.07-2.18</td>
</tr>
<tr>
<td></td>
<td>Sum of switches</td>
<td>16.47</td>
<td>6.22</td>
<td>4-27</td>
<td>28.97</td>
<td>6.64</td>
<td>16-42</td>
</tr>
<tr>
<td></td>
<td>Sum of clusters</td>
<td>6.47</td>
<td>2.65</td>
<td>1-13</td>
<td>12.40</td>
<td>4.83</td>
<td>4-26</td>
</tr>
<tr>
<td>Semantic</td>
<td>Total 3 categories</td>
<td>39.63</td>
<td>11.07</td>
<td>17-70</td>
<td>64.03</td>
<td>11.20</td>
<td>44-88</td>
</tr>
<tr>
<td></td>
<td>Mean cluster size</td>
<td>1.58</td>
<td>.32</td>
<td>.97-2.22</td>
<td>1.80</td>
<td>.34</td>
<td>1.28-2.45</td>
</tr>
<tr>
<td></td>
<td>Sum of switches</td>
<td>20.87</td>
<td>5.91</td>
<td>8-33</td>
<td>33.60</td>
<td>4.80</td>
<td>26-41</td>
</tr>
<tr>
<td></td>
<td>Sum of clusters</td>
<td>11.50</td>
<td>3.79</td>
<td>5-22</td>
<td>18.30</td>
<td>3.81</td>
<td>11-25</td>
</tr>
</tbody>
</table>

Note: Results are presented for all three letters together on the phonemic task and for all three categories together on the semantic task.
Table 3

Z-scores of eight fluency variables relative to mean control group

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>TBI Z-score</th>
<th>Control Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonemic</td>
<td>Total 3 letters</td>
<td>-1.84</td>
<td>.77</td>
<td>(-3.41)(-0.51)</td>
<td>N = 21</td>
<td>(-1.73)(2.10)</td>
</tr>
<tr>
<td></td>
<td>Mean cluster size</td>
<td>-.67</td>
<td>1.41</td>
<td>(-2.90)(4.00)</td>
<td>N = 9</td>
<td>(-1.53)(2.30)</td>
</tr>
<tr>
<td></td>
<td>Sum of switches</td>
<td>-1.91</td>
<td>.95</td>
<td>(-3.82)(-0.30)</td>
<td>N = 21</td>
<td>(-1.99)(2.00)</td>
</tr>
<tr>
<td></td>
<td>Sum of clusters</td>
<td>-1.25</td>
<td>.56</td>
<td>(-2.40)(0.13)</td>
<td>N = 12</td>
<td>(-1.77)(2.86)</td>
</tr>
<tr>
<td>semantic</td>
<td>Total 3 categories</td>
<td>-2.22</td>
<td>1.01</td>
<td>(-4.27)(0.54)</td>
<td>N = 26</td>
<td>(-1.82)(2.18)</td>
</tr>
<tr>
<td></td>
<td>Mean cluster size</td>
<td>-.67</td>
<td>.97</td>
<td>(-2.52)(1.27)</td>
<td>N = 6</td>
<td>(-1.59)(1.97)</td>
</tr>
<tr>
<td></td>
<td>Sum of switches</td>
<td>-2.70</td>
<td>1.25</td>
<td>(-5.42)(-0.13)</td>
<td>N = 25</td>
<td>(-1.61)(1.57)</td>
</tr>
<tr>
<td></td>
<td>Sum of clusters</td>
<td>-1.82</td>
<td>1.01</td>
<td>(-3.56)(0.99)</td>
<td>N = 18</td>
<td>(-1.95)(1.79)</td>
</tr>
</tbody>
</table>

Note: Z-scores were calculated on the basis of the mean and standard deviation of the raw data in the control group. Thus, the mean of z-scores within the control group was 0 for all variables and the standard deviation of z-scores was always 1 in that group.
Figure captions

Figure 1. Means of z-scores in TBI group relative to control group in each fluency variable, by test