Analysis of word clustering in verbal fluency of school-aged children

Rinat Koren, Ora Kofman *, Andrea Berger

Department of Behavioral Sciences and Zlotowski Centre for Neurosciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva 84105, Israel

Accepted 30 June 2005

Abstract

The use of clustering strategies in semantic and phonological fluency was tested in children in third (age 8–9) and fifth (age 10–11) grades. To analyze the development of clustering strategies in the fluency task, the latency to the first word and intervals between clustered and non-clustered words were recorded. Semantic fluency was greater than phonological fluency in both age groups. Children in fifth grade had greater semantic and phonological fluency than children in the third grade, concomitant with an increase in number of clusters, but not in cluster size. The greater facility of the children with semantic fluency was confirmed by a shorter latency to the first word, and significantly shorter intervals between clustered words compared to non-clustered words. Girls had shorter latencies to the first word and made more phonological clusters in the semantic task than boys. The increase in cluster number and concomitant increase in fluency in older children might be related to the development of cognitive flexibility. These data suggest that measurement of clustering strategies in the verbal fluency task can be used to assess executive function deficits in children with acquired or developmental neurological impairment or attention deficits.

Keywords: Semantic fluency; Phonological fluency; Clusters; Sex differences

The verbal fluency (VF) task is a measure of cognitive flexibility and search strategy often used to estimate cognitive functions after brain injury. Two types of fluency tasks are commonly used, semantic and phonological fluency (also known as letter fluency), each of which is considered to be related to different processes. When performance of adults was compared on phonological and semantic fluency tasks, more words were produced in semantic than in

* Corresponding author. Tel.: +972 8 647 2032; fax: +972 8 647 2088.
E-mail address: kofman@bgu.ac.il (O. Kofman).
phonological fluency (Spreen & Strauss, 1998). The VF task is dependent upon a number of different executive functions (Bolla, Lindgren, Bonaccorsy, & Bleecker, 1990; Ruff, Light, Parker, & Levin, 1997; Stuss et al., 1998) which develop during the elementary school years and attain adult levels in adolescence (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001). In children, verbal fluency output increased between age 6 and 11, with adult levels of verbal fluency being attained at age 11–12 (Anderson et al., 2001; Sauzén, Lestage, Raboutet, N’Kou, & Claverie, 2004; Temple, 1997). Six year-olds generally had very low output (<5 words/letter per minute) on the phonological or letter fluency task (reviewed in Spreen & Strauss, 1998). Previous studies have shown that processes underlying the performance of verbal fluency tasks in children are similar to those in adults, as suggested by studies of fluency in children with brain injuries (Levin, Song, Ewing-Cobbs, Chapman, & Mendelsohn, 2001), and a neuroimaging study in healthy children, which showed a similar, but more widespread activation pattern to that of adults (Guillard et al., 2000). Factor analysis of children’s performance distinguished between phonological and semantic fluency, similar to adults (Halperin, Healey, Zeitchik, Ludman, & Weinstein, 1989; Riva, Nichelli, & Devoti, 2000).

Analysis of verbal fluency production by breaking the list into clusters of words that share similar properties can shed light on the ability of the participant to recall associated words and to switch to new categories when necessary. Because the clustering reflects both memory and organizational abilities, analysis of the cluster formation has been applied to various healthy and neurologically or psychiatrically impaired adult patients in order to further understand the nature of the cognitive impairment (Elvevag, Fisher, Gurd, & Goldberg, 2002; Ho et al., 2002; Raskin & Rearick, 1996; Troyer, Moscovitch, & Winocur, 1997; Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998). A deficit in executive function in adults with brain injury often resembles performance in children who have not yet developed that particular executive ability. Thus, typically developing children, who do not have well-developed organizational skills, would be expected to be less proficient in the switching categories, similarly to adults with dorsolateral prefrontal cortical damage. Despite the potential of this analysis as a measure of internally generated cognitive flexibility in children, it is not generally used in studies with healthy children or children who have suffered brain injury. While preparing this paper, we found one other recent study on clustering in verbal fluency in children in French, which was ongoing in parallel with the current study (Sauzén et al., 2004).

Clustering of words is a strategy that has been described for episodic memory retrieval, word association and fluency tasks (Gruenwald & Lockhead, 1980; Raskin, Shiwiński, & Borod, 1992; Tröster et al., 1998; Troyer et al., 1997; Troyer, Moscovitch, Winocur, Alexander, et al., 1998; Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998; Wixted & Rohrer, 1994). The clustering process entails searching sequentially for categories of associated words, retrieving items from the category and then switching to a new category, when retrieval from the previous category fails. Semantic clusters consist of words with related meaning. Phonological clusters consist of words starting with the same letter/sound or words that rhyme. The two types of clusters, phonological and semantic, are found in both types of fluency (Laine & Niemi, 1988), and are accordingly defined as task-congruent or task-discrepant. It has been argued that semantic clusters are more automatic, relying on common rules of categorization (e.g., jungle animals, farm animals, pets), whereas phonological clustering is more laborious and relies on frontal lobe function (Ho et al., 2002). The intervals between words within clusters in
the semantic fluency task, were shorter than the intervals between words in different clusters. Moreover, when tested over a 15-min period, the clusters were separated by pauses (Gruenwald & Lockhead, 1980; Wixted & Roher, 1994).

Troyer et al. (1997) suggested that optimal fluency relies on two abilities: (a) the ability to create clusters and (b) the ability to shift between clusters. The ability to create clusters, including production of words belonging to subcategories (semantic or phonological), depends on verbal memory and word storage, which are mediated by the temporal lobe. The ability to shift between clusters, or switching, involves cognitive processes such as strategic search, cognitive flexibility and set shifting, which are mediated by the frontal lobe. They further claimed that semantic and executive abilities are equally important to semantic fluency. In the phonological task, in which few semantic clues are available (e.g., no a priori organization of words starting with a particular phoneme), there is more reliance on executive search strategies. On the other hand, Mayr (2002) argued that since there is no evidence for a trade-off between cluster size and switches, Troyer’s explanation is unfounded. He suggests that the executive component, which includes updating, stopping and initiating retrieval, is equivalent for within-cluster retrieval and switching between clusters; however, he did not relate to the phonological task, in which semantic processes make a minor contribution.

The method of analysis of cluster formation is not standardized, as various definitions of clusters have been reported. Troyer et al.’s method (1997) calculated switches and mean cluster size (number of words in a cluster minus one). Single words were designated as a cluster of size 0 and were included in the calculations. This method has been criticized because averaging a list (e.g., 12 items) that includes a single long cluster (e.g., 9 words) with three single words (cluster size 0), would produce the same mean cluster size (2) and number of switches (3), as a list that is composed of four three-word clusters (Abwender, Swan, Bowerman, & Connolly, 2001). On the other hand, Raskin’s group counted pairs of associated words, without regard to organization into longer clusters. According to this method, the number of clusters, rather than the number of switches, is used to indicate cognitive flexibility (Elvevag et al., 2002; Raskin & Rearick, 1996; Raskin et al., 1992). However, if only pairs are counted, information on cluster size, an indicator of retrieval ability, is not available. In contrast to Troyer’s method, Raskin’s method does not enumerate single words in clusters, but does relate to the ratio of total words to number of clusters. Robert et al. (1998) measured clusters as three associated words in the semantic task and two associated words in the phonological task.

The method for analysis in the present study was chosen to optimize the ability to quantify both flexibility and retrieval strategies, and combined elements of both Raskin’s and Troyer’s methods. Clusters were defined as 2 or more related words, but single words were not counted as clusters, because they suggest a failure to retrieve other words from that particular category (Abwender et al., 2001). The number of clusters was used as a measure of cognitive flexibility, as in Raskin’s studies, and cluster size was used as a measure of retrieval, as in Troyer’s studies. Both task-congruent and task-discrepant clustering were assessed, since task-discrepant clustering is a better index of active strategy utilization (Abwender et al., 2001). In order to determine if the use of clustering in children facilitates production, we also compared the mean interval between words that were included in clusters and between non-clustered words, i.e., at the point of switching from one cluster to the next or between single words. This method of analysis was adopted in order to overcome the shortcomings of previous methods.
By not counting single words as clusters, we avoided the pitfall of inflating the flexibility score (switching, according to Troyer or number of clusters in our method) when, in fact, participants were laboriously groping for words and unable to utilize an associative strategy (as would be the case if many single words were used).

According to Troyer’s model, children should show an increase in cluster number with maturation of executive functions, and cluster size should be age-invariant if vocabulary differences are held constant. However, Maye’s model for semantic fluency would predict increases in both cluster size and cluster number with maturation of executive functions. In a recent study on verbal fluency task in children aged 7–16, Sauzéon et al. (2004) showed that cluster ratio (number of clusters/total number of words) did increase with age, whereas cluster size ratio (number of words per cluster/total words) decreased until ages 11–12 and then remained stable.

Sex differences in verbal fluency have not consistently been found in several studies in adults (Demakis & Harrison, 1997; Mitrushina, Boone, & D’Elia, 1999; Tombaugh, Kozak, & Rees, 1999) and pre-adolescent children (Riva et al., 2000), although women show an advantage for speed of serial phoneme repetition (Kimura, 1999). Anderson et al. (2001) reported an overall female advantage for phonological fluency in adolescent girls. Capitani et al. (1999) reported an advantage for the semantic category tools for boys, and an advantage for the category fruits for girls. Other studies in children did not report an analysis of sex differences even though boys and girls were tested (e.g., Levin et al., 2001; Sauzéon et al., 2004). Although a consistent sex difference has not been shown in the overall fluency production, neuroimaging data suggest that adult men and women may use different strategies in word production. Slightly different patterns of brain deactivation were found, while silently performing the verbal fluency task. Women showed increased activity in the left dorsal parietal and right orbitofrontal cortex, while men showed deactivation bilaterally in the superior temporal gyrus. However, both sexes showed the basic pattern of increased activation in the left dorsolateral prefrontal cortex and right cerebellum (Schlosser et al., 1998). Women show more bilateral activation than men when performing a phonological matching task (Shaywitz et al., 1995).

In the present study, analysis of cluster formation and timing of word production was applied to two age groups of children in order to determine if they can be used to measure the development of executive processes in the VF task in children. Third and fifth grade children were chosen because, (a) they would already be proficient in reading and have output that is large enough to group into clusters, (b) they are within the prepubertal age range which shows a steady improvement in VF. Since there are no normalized data and no standard form of testing VF in Hebrew, this study was a pioneering analysis ofVF strategies in Hebrew-speaking children. Therefore, we compared a larger number of semantic categories, both common and uncommon, and sounds than are commonly used in VF studies. The hypotheses, summarized below, relate to the effect of age on fluency and organizational strategies (Hypothesis 1 and Hypothesis 2) and to the effect of sex on different organizational strategies and fluency (Hypothesis 3). Hypothesis 4 refers to the effect of category familiarity on fluency and Hypothesis 5 refers to the effect of organizational strategies on the pace of word production.

Hypothesis 1. We predicted an increase in fluency and in cluster number, but not cluster size with age.
Hypothesis 2. Task-discrepant clusters are evidence of more advanced organizational strategies (Abwender et al., 2001) and would therefore be predicted to increase with age in children.

Hypothesis 3. We hypothesized that while there would be no sex difference in overall VF production, girls would show greater proficiency for the phonological fluency strategy which might be revealed in a greater number of phonological clusters in both the semantic and phonological fluency tasks.

Hypothesis 4. It was hypothesized that uncommon categories would yield smaller total fluency. Therefore, the analysis of time intervals was conducted on the more standard semantic categories only.

Hypothesis 5. Facilitation of fluency by clustering the words would be reflected in shorter intervals between words within clusters compared to the intervals between clusters or between isolated words in both age groups.

1. Method

1.1. Participants

Participants were 72 school-aged children from two age groups: 36 third graders (17 girls)—mean age 9.03 ± 0.44, and 36 fifth graders (21 girls)—mean age 10.89 ± 0.41. All the children were native Hebrew speakers with no known diagnosis of learning disabilities. The age of the boys and girls was not significantly different in both age groups (third grade: \( F(1,34) = 0.06, \text{M.S.E.} = 0.19 \), fifth grade: \( F(1,34) = 1.79, \text{M.S.E.} = 0.16 \)). The children were recruited from two schools, two third grade classrooms per school, and one fifth grade classroom in one school and two in the other school. Both schools were located in similar urban neighborhoods, thus ensuring that the general composition of the population was homogeneous and not from the more affluent suburban population. All children tested were included in the analysis.

1.2. Procedure

The participants were tested on the vocabulary subtest of the WISC-R (Wechsler, 1991, Hebrew version) to ensure that they all performed within the normal range for their age. Since vocabulary can be an important variable in determining fluency (Ruff et al., 1997), this variable was used as a covariate in the analysis when it was significantly correlated with the dependent measure. The range of the normalized vocabulary scores was identical for both grades (7–14). The age-normalized means and standard deviations were 11.44 ± 1.98 for third grade and 10.94 ± 1.69 for fifth grade, i.e., the children in each grade had average vocabulary levels for their age group.

In addition, handedness was scored using a modified version of the Edinburgh handedness inventory test (Oldfield, 1971 translated into Hebrew). However, the small number of left-handed children (three boys and two girls) did not allow us to analyze handedness as a variable.
An ANCOVA for total fluency in both tasks, omitting the five left-handed children showed similar results to the fluency analysis when they were included. Therefore, the five left-handed children were included in all further analyses.

Following these tests, the semantic and phonological fluency tasks were tested as described below. Semantic fluency was measured in the following categories: ‘animals’, ‘food’, ‘clothes’ and ‘things that are seen on the street’ (designated below as ‘street’). ‘Animals’ and ‘food’ are categories used in the most studies on fluency (Hurks et al., 2004; Mitrushina et al., 1999; Riva et al., 2000; Tombaugh et al., 1999). ‘Clothes’ and ‘street’ were chosen as uncommon categories, in order to examine the effect of category familiarity on fluency. The phonological categories were the Hebrew letters $\lambda$, $\alpha$, $\beta$, $\gamma$, $\wp$ (g, d, p, r, sh, respectively). These letters were chosen because they do not form phonemes that can be used in the conjugation of verbs and because the phone represented by each of the letters is unique to that letter (in contrast to F and S in English which have the same sounds as ph or a soft C, respectively). In Hebrew, verbs are conjugated by adding phonemes to the beginning and end of a root word. In addition, articles and conjunctions (e.g., the, which, and) are formed by a single letter at the beginning of the word. Letters used for conjugation of verbs or as articles were not used in this study.

The experimenter pronounced the phones rather than naming the letter. The order of the two fluency tasks was counterbalanced, and the order of the letters/categories in each fluency task varied among participants in a balanced order, determined by a Latin square. The children were told to produce as many words as they can, as fast as they can, without using proper names or different forms of the same words. The instructions did not refer to sub-categories (e.g., domestic animals, jungle animals) and did not give examples for the categories used in the test. The instructions included an example of the category ‘furniture’ and the phone ‘m’ for semantic and phonological fluency, respectively. The experimenter gave three to four examples from each of these categories (which were not part of the test) and asked the child to provide further examples, thus verifying that the child understood the instructions before proceeding. After presenting the category, fluency was recorded for 60 s, after which the child was told to stop.

Fluency for each category was the total number of words produced, excluding errors (e.g., repetitions, proper names etc.). Total fluency for each fluency type (semantic/phonological), was calculated as the mean fluency produced in the related categories.

1.3. Clusters—number and size

The words produced by each child, including errors, were categorized as belonging to clusters or as non-clustered words. Number of clusters and mean cluster size were measured for each category/phone. Single words were not counted as clusters. In order to examine organizational strategies, our measures focused on clusters of at least two words. Furthermore, in order to determine how the organization affects timing of word production, the mean intervals between uttering clustered words and non-clustered words were compared.

A semantic cluster was defined as two or more successive words belonging to a specific subcategory. The subcategories were either determined a priori, or were defined after analyzing the typical responses made by the children. While there is no standard method to
sort semantic clusters, there is an underlying assumption that semantic categories are firmly established.

There were two slightly different definitions of phonological clusters, according to the type of fluency. In the phonological fluency task, the phonological cluster was defined as two or more successive words corresponding to the following rules: (1) words beginning with the same syllable (consonant + vowel), (2) words beginning with two identical consonants and (3) words that rhyme. In the semantic fluency task, in addition to above-mentioned criteria, words beginning with the same consonant were included. This criterion obviously could not be considered in the phonological task since, by definition, all words begin with the same letter.

1.4. Analysis of time intervals

All responses were recorded on a cassette recorder, and three time measurements were derived from the fluency tasks using a digital stopwatch. The interval to the first word was measured from the time the experimenter said, “go” until the child uttered its first response. Following the analysis of words into clusters, the intervals between words within the clusters (clustered words), and the intervals between the clusters and between single words (non-clustered words) were measured for the ‘animals’ and ‘food’ semantic categories and for all five phones. Only the two most common categories were used for the semantic fluency task, because our data (below) showed greater fluency and organization for these common categories, as opposed to the uncommon categories of clothes and street.

Because the quality of some of the recordings was too poor for accurate analysis, and some of the words uttered by the children were inaudible, the number of participants in the timed measures is smaller than those included in the fluency task. Different phones or categories were inaudible for different children; however, because of the analysis involved repeated measures, omission of any part of the task for a particular participant resulted in elimination of this participant from the analysis. Despite this limitation, and because the final distribution of participants between grade and sex was balanced, it was decided to analyze the time intervals. This type of analysis has not yet been reported in this type of task in children. For semantic fluency, analysis was done for 13 boys and 13 girls in the third grade, and 10 boys and 12 girls in the fifth grade. For phonological fluency, the number of clear recordings included 13 boys and 14 girls in third grade and 9 boys and 12 girls in fifth grade.

1.5. Statistics

Analysis of variance (ANOVA) was used to test the effect of grade, sex and categories on the measurements described above, and analysis of covariance (ANCOVA) was used when considering the additional influence of vocabulary. According to the results, additional analyses were made. Pearson correlations were used to assess the relationship of cluster number and size to fluency. We also performed a hierarchical regression analysis to identify the influence of each variable on fluency. Analyses were done with Statistica software (Statsoft Inc., version 6.1), except for the multiple regression, which was done with SPSS.
2. Results

A two-way ANCOVA for the effect of task (semantic, phonological) as a repeated measure and grade (third, fifth), with vocabulary as a covariate was conducted on the total fluency (number of words) scores. In subsequent sections, each task was further analyzed by category. The covariate vocabulary was significant, $F(1,67) = 6.99$, M.S.E. = 6.13, $p < .01$, and was therefore included in subsequent analyses, only if it significantly correlated with the measure. The ANCOVA further revealed that total semantic fluency was significantly greater than total phonological fluency $F(1,67) = 17.96$, M.S.E. = 3.03, $p < .0001$, and overall the fifth grade children had greater fluency than third grade children $F(1,67) = 18.95$, M.S.E. = 6.14, $p < .0001$, as predicted by Hypothesis 1. In addition, the interaction between grade and task was significant $F(1,67) = 9.74$, M.S.E. = 3.03, $p < .005$ (Fig. 1). Subsequent analysis of the interaction showed that, as predicted by Hypothesis 1, there was a significant increase in both total semantic $F(1,67) = 22.17$, M.S.E. = 6.15, $p < .00001$, and total phonological fluency $F(1,67) = 4.93$, M.S.E. = 3.03, $p < .05$, between third and fifth grades, although the increase in semantic fluency was larger.

2.1. Semantic fluency by category

Semantic fluency was analyzed by three-way ANCOVA for the effects of grade and sex as between-participant variables, category as a repeated variable (‘animals’, ‘food’, ‘clothes’ and ‘street’) and vocabulary as a covariate. Main effects were found for grade, showing greater fluency for the fifth graders, $F(1,66) = 22.17$, M.S.E. = 24.62, $p < .001$ (Hypothesis 1) (Table 1), and category $F(3,198) = 2.63$, M.S.E. = 7.27, $p = .05$ (Table 2). Although the difference between the categories was only marginally significant, a post hoc Scheffé test was conducted in order to determine if there was a significant difference between common and uncommon categories. The post hoc analysis showed higher fluency in the categories ‘food’ and ‘animals’ compared

![Fig. 1. Mean ± S.D. of semantic and phonological fluency for children in third and fifth grade. Semantic fluency > phonological fluency ($p < .001$). # Fifth grade pupils > third grade pupils ($p < .01$). ★★ Task × grade interaction ($p < .005$).]
Table 1
Mean (S.D.) of total fluency, number of clusters and cluster size by grade and task

<table>
<thead>
<tr>
<th>Grade</th>
<th>Semantic fluency</th>
<th>Phonological fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total fluency</td>
<td>Number semantic</td>
</tr>
<tr>
<td></td>
<td>clusters</td>
<td>clusters</td>
</tr>
<tr>
<td>3</td>
<td>11.00 (2.26)</td>
<td>3.42 ** (0.87)</td>
</tr>
<tr>
<td>5</td>
<td>13.72 (2.78)</td>
<td>4.40 (1.27)</td>
</tr>
</tbody>
</table>

* Comparison between grades: p < .001.
** Comparison between grades: p < .005.
# Comparison between grades: p < .05.

to the other categories (p < .001), as predicted by Hypothesis 4. No main effect for sex and no interactions between the independent variables were found. Because the correlation between the vocabulary and the semantic fluency was not significant (r = .12), vocabulary was not used as a covariate in the subsequent analysis of the clusters.

Three-way ANOVA with repeated measures tested the effect of grade, sex (between participants) and category (repeated measure) on the mean semantic cluster size. No main effect for grade or sex was found, but there was a main effect for category $F(3,201)=17.83$, M.S.E. = 0.38, p < .001 (Table 2). Significantly larger clusters were made in the category ‘animals’ compared to the category ‘clothes’ (p < .05). Furthermore, there was a significant interaction between category and grade, $F(3,201)=6.3$, M.S.E. = 0.38, p < .001. Further analysis of the interaction using a post hoc Scheffé test revealed significantly larger clusters in the category ‘food’ compared to the other categories in the fifth grade only (p < .005, Mean (S.D.) = 3.36 (1.13)).

2.2. Phonological fluency by category

The effect of grade and sex, as between participant variables, and category (phone), as a repeated measure, on total phonological fluency was analyzed by three-way ANCOVA with the vocabulary score as a covariate. As expected, we found a main effect of grade, showing greater fluency in fifth grade children compared to third graders, $F(1,67)=4.74$, M.S.E. = 3.0,

Table 2
Mean (S.D.) of semantic fluency and number and size of semantic clusters, according to categories, for children of both grades 3 and 5

<table>
<thead>
<tr>
<th>Category</th>
<th>Total fluency</th>
<th>Number of semantic clusters</th>
<th>Size of semantic clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>14.80 (4.18)</td>
<td>5.55 (1.95)</td>
<td>2.64 (0.44)</td>
</tr>
<tr>
<td>Street</td>
<td>11.25 (3.31)</td>
<td>2.92 (1.77)</td>
<td>2.43 (0.66)</td>
</tr>
<tr>
<td>Food</td>
<td>12.93 (4.28)</td>
<td>4.37 (1.93)</td>
<td>3.02 (0.91)</td>
</tr>
<tr>
<td>Clothes</td>
<td>10.46 (2.84)</td>
<td>2.80 (1.23)</td>
<td>2.33 (0.62)</td>
</tr>
</tbody>
</table>

* p < .001 compared to all of the other categories.
" p < .005 compared to all of the other categories, for fifth grade only.
* p < .05 ‘animals’ compared to ‘clothes’.
p < .05, as predicted by Hypothesis 1 (Table 2). Vocabulary was significantly correlated with phonological fluency ($r = .23$, $p < .05$). No main effects or interactions were found for sex or category. Since one of our goals was to determine which phones in Hebrew would yield higher fluency, we also conducted an ANOVA for the effect of sex and grade as between participant variables and category (phone) as a repeated measure. A main effect of category was found, $F(4,272) = 15.53$, M.S.E. = 2.97, $p < .00001$. Post hoc Scheffé test showed greater fluency for the phones $g$, $sh$ ($p < .005$). The mean (S.D.) for each phone was $S_tg = 6.60 (2.64)$, $G = 6.47 (2.47)$, $P = 5.35 (2.27)$, and $R = 4.82 (2.06)$.

2.3. Semantic clusters

A three-way ANOVA with repeated measures was made to test the effect of grade, sex (between participant) and category (repeated measure) on the mean number of semantic clusters. As predicted in Hypothesis 2, there was a main effect for grade $F(1,67) = 13.16$, M.S.E. = 4.82, $p < .005$ (Table 1), showing that fifth grade children produced more clusters than third grade children, and a main effect for category $F(3,201) = 53.69$, M.S.E. = 2.15, $p < .001$ (Table 2). The post hoc Scheffé test showed that significantly more clusters were produced in the categories ‘food’ and ‘animals’ compared to the other categories ($p < .001$), in accord with Hypothesis 4. No main effect for sex and no interactions were found.

2.4. Phonological clusters

A three-way ANCOVA for the effects of grade and sex, as between participant variables, category as a repeated measure, and vocabulary as a covariate was conducted for the number of clusters and for cluster size. No main effects and no interactions were found.

2.5. Relationship of cluster number and size to verbal fluency

High positive Pearson correlations were found between total fluency and number of clusters (semantic: $r = .85$, phonological: $r = .87$, $p < .05$), and low positive correlations were found between total fluency and mean cluster size (semantic: $r = .34$, phonological: $r = .46$, $p < .05$). The difference between these correlations in each fluency task was significant (semantic $p < .001$, phonological $p < .03$). Cluster size and number were significantly correlated for semantic ($r = .30$, $p < .05$) but not for phonological fluency ($r = .04$).

2.6. Multiple regression

A hierarchical regression was used to assess the influence of the number of clusters and the cluster size on fluency. In the first step, age and vocabulary were entered to control for these variables. For both types of fluency, the second step of the analysis showed that the number of clusters, but not the cluster size, had a significant contribution above and beyond the effect of the variables entered in the first step (see full results in Table 3). As can be seen, the models predict 74% of the variance of the dependent variable in the semantic fluency and 49% of the variance in the phonological fluency.
Table 3
Hierarchical regression full results for each fluency task

<table>
<thead>
<tr>
<th>Variables</th>
<th>β</th>
<th>ANOVA for model</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semantic fluency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.49**</td>
<td>F(2,68) = 11.5**</td>
<td>0.23</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.17*</td>
<td>F(4,66) = 49.8**</td>
<td>0.74</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of clusters</td>
<td>0.74**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of clusters</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phonological fluency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.32**</td>
<td>F(2,69) = 8.8**</td>
<td>0.18</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.35**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.25**</td>
<td>F(4,67) = 18.57**</td>
<td>0.49</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of clusters</td>
<td>0.54**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of clusters</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .01.

**p < .001.

2.7. Phonological clusters in the semantic fluency task

The effects of grade and sex on total number of phonological clusters in the semantic fluency task (task discrepant clustering) were analyzed by two-way ANOVA. A main effect for grade was found, showing more phonological clusters made by fifth grade children, F(1,68) = 4.93, M.S.E. = 4.41, p < .05, in accord with Hypothesis 2. The means (S.D.) were third grade 3.00 (1.81), fifth grade 4.65 (2.43).

There was also a main effect for sex, showing more phonological clusters made by girls F(1,68) = 6.31, M.S.E. = 4.41, p < .05, as predicted by Hypothesis 3. The group means (S.D.) were: girls 4.44 (2.31), boys 3.13 (1.98).

2.8. Response time to the first word

A three-way ANOVA was made to test the effect of grade and sex as between participant variables and task (phonological versus semantic fluency) as a repeated measure, on the mean response time to the first word. There was a main effect of sex, with a significantly shorter response time for girls compared to that of boys F(1,68) = 7.16, M.S.E. = 4.76, p < .01, mean (S.D.) = 2.20 (1.23) and 3.23 (1.84) s, respectively. This was in agreement with Hypothesis 3 that suggested that females would show greater facility in the VF tasks.

In addition, there was a significant difference between the two types of fluency tasks. The response time in the semantic fluency task was significantly shorter than in the phonological flu-
The mean (S.D.) of intervals (s) between words in the semantic fluency task is presented in Table 4. The duration of the interval between words within clusters (clustered) and intervals between clusters or between single words that were not included in a cluster (non-clustered) words was analyzed using ANOVA with category (‘animal’ versus ‘food’) and cluster (non-clustered versus clustered) as repeated measures and grade and sex as the between participant variables after the Levene’s test showed no significant differences for any of these variables. A significant main effect for cluster was found, \( F(1,44) = 47.87, \text{M.S.E.} = 12.98, p < .00001 \), such that the interval between clustered words was significantly shorter than the interval between non-clustered words. The main effect of category, \( F(1,14) = 6.91, \text{M.S.E.} = 7.43, p < .05 \), revealed that the mean interval between words in the ‘food’ category was shorter than the mean interval in the ‘animal’ category. The main effect of grade, \( F(1,44) = 4.97, \text{M.S.E.} = 11.36, p < .05 \), indicated that fifth graders had shorter intervals than third graders, which supports the fifth hypothesis, as well as the second hypotheses suggesting better organizational strategies in older children.

Moreover, a four-way interaction between grade \( \times \) sex \( \times \) category \( \times \) cluster, \( F(1,44) = 6.31, \text{M.S.E.} = 9.24, p < .05 \), was found and analyzed further by separate ANOVAs for each category. The three way ANOVA (cluster \( \times \) sex \( \times \) grade) was significant for ‘animals’ \( F(1,46) = 6.57, \text{M.S.E.} = 7.26, p < .05 \), but not for ‘food’ \( F(1,45) = 1.03, \text{M.S.E.} = 14.49 \). Thus, when naming food, there was a main effect, such that the intervals between clustered words were shorter than intervals between non-clustered words, \( F(1,45) = 32.69, \text{M.S.E.} = 14.49, p < .001 \). No significant interactions were found. In the ‘animals’ category, only the third grade girls showed a significant difference between intervals of non-clustered and clustered words (post hoc Scheffé tests between non-clustered and clustered words, \( p < .01 \)) (Table 4). Due to the small number of phonological clusters in the semantic fluency task, the intervals between phonologically clustered and non-clustered words were not analyzed.
The intervals in seconds between phonologically clustered and non-clustered words in phonological fluency were analyzed in a four-way ANOVA for the five phones, with phone and cluster as repeated variables and grade and sex as between-participant variables. A significant effect was found for phone \(F(4,64) = 7.73, \text{M.S.E.} = 24.43, p < .001\). Post hoc Scheffé analysis revealed that the intervals between words for the phone \(sh\) were significantly shorter than the intervals for the phone \(r\). No main effects or interactions with sex were found. In order to derive an overall measure of the intervals per task and grade, the mean of the between- and within-cluster intervals was calculated for each participant, and a \(2 \times 2 \times 2\) ANOVA was performed on this measure for the effects of sex and grade, with cluster as a repeated measure. A significant difference was found between grades, such that fifth grade children had shorter intervals overall, \(F(1,44) = 8.16, \text{M.S.E.} = 13.41, p < .01\), supporting the Hypotheses 2 and 5.

In other words, shorter time intervals were associated with better organizational strategies (clusters) and both were seen more in older children. In addition, there were significantly shorter intervals between clustered words than between non-clustered words, \(F(1,44) = 7.11, \text{M.S.E.} = 13.19, p < .05\). In the third grade the means (S.D.) were 9.63 (3.49) s for intervals between clustered words and 11.74 (4.53) s for intervals between words not in clusters. In the fifth grade the mean (S.D.) interval between clustered words was 7.66 (3.00) s and between non-clustered words it was 9.29 (3.05) s.

3. Discussion

The results in the present study showed that: (a) semantic fluency increased more with age than phonological fluency and that (b) older children had a larger number of semantic clusters, but no difference in cluster size compared to younger children. The age range in this study, albeit restricted, includes children whose verbal fluency should not have attained adult levels. Executive functions, such as set maintenance and impulse control, have been shown to mature fully by age 10; however verbal fluency and other complex planning tasks did not reach adult levels until age 12 (Temple, 1997). The semantic fluency in Hebrew as measured by the categories ‘food’ and ‘animals’, and the phonological fluency in the present study in Hebrew are similar to results in English (Halperin et al., 1989), Italian (Riva et al., 2000) French (Sauzéon et al., 2004) and Dutch (Hurks et al., 2004) from children of this age, i.e., approximately 11–15 words for semantic fluency and approximately 6–8 words per letter for phonological fluency. The study by Sauzéon et al. (2004), which is the only other study to examine clustering strategies, albeit, using a slightly different methodology in French, found similar results. There were more task-congruent clusters than task-discrepant clusters, an increase in cluster size number with age, and no difference in cluster size ratio between similar age groups for the phonological fluency task.

The maturational change in cluster number with no change in cluster size supports the model proposed by the studies in adults with brain damage, suggesting that the age difference in performance can be accounted for mainly by maturation of retrieval strategies that are likely related to frontal lobe maturation. Patient studies have shown that the prefrontal cortex, especially the dorsolateral area, is particularly critical for phonological fluency (Baldo & Shimamura, 1998; Martin, Wiggs, Lalonde, & Mack, 1994; Stuss et al., 1998), while semantic fluency is
more affected by temporal lobe lesions (Gaillard et al., 2000; Martin et al., 1994). Cluster analysis of VF in patients with lesions in frontal areas (Troyer, Moscovitch, Winocur, Alexander, et al., 1998), dementia associated with neurodegenerative disease (Ho et al., 2002; Raskin et al., 1992; Tröster et al., 1998), and mild traumatic brain injury (Raskin & Rearick, 1996) suggested that the number of clusters is dependent on the frontal lobe, and cluster size is dependent on the temporal lobe (Stuss et al., 1998; Tröster et al., 1998; Troyer et al., 1997; Troyer, Moscovitch, Winocur, Alexander, et al., 1998).

Frontal lobe maturation continues throughout childhood and adolescence, as indicated by the development of white matter and gray matter (Giedd et al., 1999), and elimination of synapses (Huttenlocher & Dabholkar, 1997). Although the link between maturation of fluency processes and frontal lobe maturation is tenuous, it is supported by studies of children with brain injuries. Left frontal lesions in older children (9–13) affect the VF performance more adversely than equivalent lesions at a younger age (7 years) (Levin et al., 2001). This suggests a more established functional commitment of the left frontal lobe to the VF task at late childhood and adolescence.

Mayr (2002) claimed that the executive component is essential for updating and retrieval within a cluster, and that it is invariant for each word retrieved in the semantic fluency task. Accordingly, one would predict an age difference in cluster size and in errors of monitoring (repeated words, self corrections). There were too few errors to conduct parametric statistical analysis; however, no significant age differences were found in monitoring errors between fifth and third grade children (Koren, 2002). Although cluster size and number were positively correlated to total fluency, cluster number had a more prominent contribution. This suggests that even if cluster size and number measure the same process, cluster number is a more sensitive index of age-related changes in VF in school-aged children.

In both types of fluency the positive correlations between the fluency size and the number of clusters were significantly higher than the correlations between the fluency size and cluster size. Moreover, hierarchical regression showed that total fluency was predicted by the number of clusters, after controlling for the shared variance explained by age and vocabulary. This suggests that in the ages tested, the frontal component (number of clusters) is more important than the temporal component (cluster size). This finding contrasts with results of studies in adults showing an equal contribution of both components in semantic fluency (Troyer et al., 1997). The intervals between uttering words within clusters was significantly shorter than the interval between non-clustered words in the semantic task, confirming the notion that fluency production occurs in bursts of associated words, with more effortful switches between clusters (Elvevag et al., 2002; Gruenwald & Lockhead, 1980).

These findings have critical implications for clinical practice and research on executive functions. Since clustering appears to predict organizational strategies, use of this analysis can detect executive function problems in children with developmental neurological deficits such as ADHD or epilepsy. For example, Hurks et al. (2004) found that fluency was impaired in the first 15 s of the 1-min test in children with ADHD, but not in the period from 16 to 60 s. Analysis of the pattern of production of word clusters, as well as the intervals between words might reveal immature patterns of clustering, leading to slower production. Children might then be encouraged, through oral or computer word games to learn appropriate organizational strategies. Similarly, in a longitudinal study, children who had Rolandic epilepsy showed
persistent deficits in VF even after their memory impairments and other executive functions had recovered (Lindgren et al., 2004). In such a situation, analysis of the VF output using the methods described in this study, could provide evidence for different patterns of deficits. For instance, fewer clusters would suggest deficient organizational strategies. Generalized slowing might suggest motor deficits, whereas slowing only between clusters or single words with normal paced production of clustered words, might suggest retrieval and switching deficits. Similar strategies could be used to follow the recovery of executive functions in children following head injury or brain surgery.

In accord with our hypothesis that girls would show greater facility on the phonological task, girls made more task-discrepant, phonological clusters in the semantic fluency task and also had shorter latencies to the first word overall. The use of task-discrepant clustering (e.g., phonological clustering in the semantic task) suggests that the girls showed greater cognitive flexibility and more effortful searching (Abwender et al., 2001). Kimura (1999) reviewed studies suggesting that girls show a greater tendency to use phonological strategies compared to boys. It is likely that gonadal hormones or differences in brain organization do make subtle contributions to performance of VF tasks, since female adolescents outperformed males on a phonological fluency task (Anderson et al., 2003) and testosterone was negatively correlated with VF in elderly men (Wolf & Kirschbaum, 2002). The present findings suggest that using cluster and timing analysis in the VF task might reveal subtle performance and strategy differences between the sexes. Further research on executive functions in children with atypical hormonal status, such as congenital adrenal hyperplasia, or atypical chromosomal conditions such as XO Turner’s would be required to elaborate on these findings.

In conclusion, the present study found increased fluency and a change in organization of output with age, which was manifested in a higher number of clusters, suggesting maturation of search strategies. A more careful analysis of response patterns, such as that described for typical children in this study, allows detection of subtle maturational changes and could be used to analyze the nature of executive function deficits in children with developmental or acquired neurological impairment.

Acknowledgements

The authors thank Dr. Na’ama Atzava-Poriah and Ms. Tali Milstein, and the principals and teachers of ‘Netivei Am’ and ‘Orian’ elementary schools in Beersheva, Israel.

Appendix A

Examples of semantic (top) and phonological (bottom) fluency in a child from the fifth grade and a child from the third grade. Pronunciation follows the key in Webster’s Collegiate Dictionary (1947). Clustered words are in boldface. J – Semantic cluster } Phonological cluster
### References


Wolf, O. T., & Kirschbaum, C. (2002). Endogenous estradiol and testosterone levels are associated with cognitive performance in older women and men. Hormones and Behavior, 41, 259–266.