Proficiency and Control in Verbal Fluency Performance across the Lifespan for Monolinguals and Bilinguals

Deanna C. Friesen¹, Lin Luo¹, Gigi Luk², & Ellen Bialystok¹

1. York University
2. Harvard University

Address for Correspondence:
Ellen Bialystok
Department of Psychology
York University
4700 Keele Street
Toronto, Ontario M3J 1P3
Canada
Email: ellenb@yorku.ca

Authors’ Note
The research reported in this paper was funded by grant R01HD052523 from the US National Institutes of Health and grant A2559 from the Natural Sciences and Engineering Research Council of Canada to EB.
Abstract

The verbal fluency task is a widely used neuropsychological test of word retrieval efficiency. Both category fluency (e.g., list animals) and letter fluency (e.g., list words that begin with F) place demands on semantic memory and executive control functions. However, letter fluency places greater demands on executive control than category fluency, making this task well-suited to investigating potential bilingual advantages in word retrieval. Here we report analyses on category and letter fluency for bilinguals and monolinguals at four ages, namely, 7-year-olds, 10-year-olds, young adults, and older adults. Three main findings emerged: 1) verbal fluency performance improved from childhood to young adulthood and remained relatively stable in late adulthood; 2) beginning at 10-years-old, the executive control requirements for letter fluency were less effortful for bilinguals than monolinguals, with a robust bilingual advantage on this task emerging in adulthood; 3) an interaction among factors showed that category fluency performance was influenced by both age and vocabulary knowledge but letter fluency performance was influenced by bilingual status.

Keywords: Verbal Fluency, Bilingualism, Executive Control, Language Proficiency, Lifespan
Proficiency and Control in Verbal Fluency Performance across the Lifespan for Monolinguals and Bilinguals

Research on bilingualism has typically investigated language processing and executive control mechanisms separately. As a result, apparently dichotomous bilingual consequences have been observed relative to monolinguals, specifically, bilingual limitations on language tasks and bilingual advantages on executive control tasks (Bialystok, Craik, Green, & Gollan, 2009). Compared to monolinguals, bilinguals obtain lower scores on standard measures of English vocabulary (Bialystok & Luk, 2012; Bialystok, Luk, Peets, & Yang, 2010), experience more tip of the tongue states (Gollan & Silverberg, 2001) and are slower to name pictures in both their languages (Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008). In contrast, bilinguals outperform monolinguals on non-verbal tasks that require them to resolve conflict and switch between tasks (e.g., Costa, Hernández, & Sebastián-Gallés, 2008; Prior & MacWhinney, 2010; however, see Paap & Greenberg, 2013 for a possible exception). To understand how the linguistic aspects of bilingualism interact with nonverbal executive control, the two need to be examined together. One means of investigating this interaction is by using the verbal fluency task, which requires both language proficiency and varying levels of executive control during lexical retrieval in a language production task.

The bilingual advantage in executive control tasks is believed to arise from the constant need to manage attention to two language systems during language production and comprehension (Bialystok et al., 2009; see Hilchey & Klein, 2011 for a review of
conditions under which a bilingual executive control advantage typically occurs). A large body of research has demonstrated that both of a bilingual’s languages are concurrently activated (e.g., Friesen & Jared, 2012; Marian & Spivey, 2003; see Kroll, Dussias, Bogulski, & Valdes-Kroff, 2012, for review), so bilinguals need to recruit attentional control processes to prevent interference from the unwanted language (Green, 1998; Abutalebi & Green, 2007). This extensive practice controlling attention to two language systems during language processing may result in a more efficient executive control system; that is, fewer resources may be require to monitor or resolve conflict (Bialystok et al., 2009). Although this executive control advantage has been demonstrated in nonverbal tasks (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Costa et al., 2008), only a few studies involving bilinguals have examined the role of executive control during language production and none has done so across the lifespan. Development of executive control is protracted and extends into adolescence (Casey, Getz, & Galvan, 2008), and there is little evidence to date for how executive control processes are engaged during language production for monolingual and bilingual children during development, or how this interaction may extend across the lifespan into late adulthood.

Most bilingual studies of vocabulary size show deficits in receptive measures (e.g., Bialystok et al., 2010 for children; Bialystok & Luk, 2012 for adults), but language production involves additional processes, including executive control, and these are enhanced in bilinguals. Therefore, language production should reveal more complex performance outcomes than the usual receptive measures. Here we use verbal fluency, a widely used neuropsychological measure of lexical retrieval efficiency. Individuals are given a time constraint within which they must generate as many words as possible that
fit a criterion corresponding to either a category (e.g., items of clothing, animals) or an initial letter (e.g., words that begin with F). Since participants must employ attentional resources to access vocabulary knowledge and inhibit responses that do not fit the criterion, both conditions require semantic memory as well as executive control functions such as working memory, response inhibition, conflict monitoring and search strategies (Hurks, 2012; Martin, Wiggs, Lalonde, & Mack, 1994; McDowd et al., 2011; Rosen & Engle, 1997). Therefore, this production task necessitates the recruitment of both language knowledge and executive control functions, although the relative contributions of each are not-well specified (McDowd et al., 2011).

In both verbal fluency conditions, participants may utilize search strategies that draw on executive control during lexical access such as clustering (i.e., grouping responses based on a subcategory, e.g., types of shirts) or switching (i.e., shifting to a new subcategory), but because of the different search criterion, the executive control demands are greater in the letter condition. Although there are potentially more exemplars for a letter category than a semantic category, individuals tend to generate fewer items during letter fluency than during category fluency (e.g., Gollan, Montoya, & Werner, 2002; Kormi-Nouri, Moradi, Moradi, Akbari-Zardkhaneh, & Zahedian, 2012; however, see Azuma et al., 1997). One possible reason is that the task demands for category fluency are consistent with the structure of semantic memory; concepts are clustered along semantic properties and speakers can take advantage of this organization to generate category members (Luo, Luk, & Bialystok, 2010). In contrast, generating words based on a phonemic cue is not a common strategy in word retrieval and lexical entries are not listed alphabetically (Strauss, Sherman, & Spreen, 2006). Consistent with
this argument, clustering has been found to account for more variance in category fluency and switching accounts for more variance in letter fluency (Filippetti & Allegri, 2011). Moreover, letter fluency requires that participants inhibit naturally generated but irrelevant semantic associates in order to focus on the letter criterion. The restrictions to exclude morphological variants, proper names, and numbers further increase the monitoring demands, and therefore the involvement of executive control (Delis Kaplan, & Kramer, 2001a; Kemper & McDowd, 2008).

Both behavioral and neuroimaging research indicate that category and letter fluency differentially recruit executive control functions. Martin et al. (1994) found a double dissociation in which performing a concurrent finger tapping task disrupted letter fluency, whereas an object decision task produced interference in the category condition. The former was proposed to influence executive control functions localized in the frontal lobes and the latter recruited semantic processing in the temporal lobes. Grogan, Green, Ali, Crinion, and Price (2009) reported different structural and functional neural correlates for each of the fluency conditions. They observed that greater grey matter density and higher activation in the caudate were associated with better performance on letter fluency relative to category fluency. This neural structure has been implicated in the executive control network (e.g., Simard et al., 2011). In contrast, the left inferior temporal cortex, a region known to be important for semantic access, was associated with the difference between category and letter fluency, again indicating that greater density in this region was associated with better performance on category fluency. A similar dissociation was observed in two patients with chronic aphasia. Baldo, Schwartz, Wilkins, and Dronkers (2010) reported that a patient with a lesion in the left temporal
lobe exhibited deficits in category fluency whereas another patient whose lesion was in the left frontal lobe was impaired in letter fluency (see also Gourovitch et al., 2000; Jurado, Mataro, Verger, Bartumeus & Junque, 2000 for similar findings). These brain regions that are associated with better letter fluency have also been shown to be active during bilingual language switching (Luk, Green, Abutalebi, & Grady, 2012). Given the behavioral evidence demonstrating bilingual advantages in executive control coupled with evidence from neuroimaging studies demonstrating that the letter task differentially recruits these executive control networks, bilinguals should show an advantage on the letter task but not on the category task.

Studies that have investigated bilingual and monolingual performance in verbal fluency have used similar procedures but reported mixed results. Studies typically report that monolinguals produce more items than bilinguals during category fluency (Gollan et al., 2002; Portocarrero, Burright, & Donovick, 2007; Rosselli et al., 2000; Sandoval, Gollan, Ferreira, & Salmon, 2010), but results are more varied for letter fluency. All of the following patterns have been reported for letter fluency: better performance for monolinguals than bilinguals (Sandoval et al., 2010), equivalent performance for monolinguals and bilinguals (Kormi-Nouri et al., 2012; Portocarrero et al., 2007; Rosselli et al., 2000), and better performance for bilinguals than monolinguals (Kormi-Nouri et al., 2012; Ljungberg, Hansson, Andrés, Josefsson, & Nilsson, 2013). For example, Kormi-Nouri et al. (2012) found that Turkish-Persian bilingual children in Grade 1 outperformed both Persian monolinguals and Kurdish-Persian bilinguals on the Persian letter task, but no group differences were observed on their samples of children in Grade 2 to Grade 5. It is not clear why an advantage would emerge in a single bilingual group
and disappear at later ages. One possible reason for these divergent findings is the proficiency of the bilinguals in the language of testing. In monolingual samples, language proficiency has been shown to be positively associated with verbal fluency performance (e.g., Hedden, Lautenschlager, & Park, 2005; Salthouse, 1993) yet proficiency is rarely assessed in bilingual studies.

Bialystok, Craik, and Luk (2008) postulated that a bilingual advantage in letter fluency may be masked by weak language proficiency. That is, performance on verbal fluency depends on both the quality of the language representations in the language of testing and the executive control processes that are recruited. Thus, bilinguals and English monolinguals were assessed on English vocabulary knowledge and participants were divided into two groups corresponding to low and high English proficiency based on a median-split of their scores. By employing an objective measure of proficiency in the test language, issues concerning whether the native language is also the dominant language can be circumvented. The high proficiency bilinguals and monolinguals did not differ on English proficiency, whereas the low proficiency bilinguals had significantly lower English vocabulary scores than the other two groups. For the category fluency condition, monolinguals and high proficiency bilinguals generated an equivalent number of words, and both of these groups produced significantly more words than did the low proficiency bilingual group. For the letter fluency condition, the low proficiency bilingual group produced an equivalent number of words as the monolinguals, but the high proficiency bilingual group produced significantly more words than either of those two groups. Thus, both bilingual groups demonstrated a relative increase in performance on the letter fluency task compared to the monolinguals; for the low proficiency bilingual
group, performance in letter fluency brought them up to the level of the monolinguals, suggesting relatively equal but opposite effects of poorer vocabulary knowledge and better executive control ability. The high proficiency group whose category fluency performance was comparable to the monolinguals went on to surpass the monolinguals in the letter fluency task. The claim is that these increases in letter fluency performance by bilinguals, irrespective of their performance on category fluency, is attributable to better executive control that is required for this task.

Luo et al., (2010) conducted a follow-up study with a new group of young adults. The results replicated the findings reported in Bialystok et al. (2008) in terms of the number of items produced in letter and category fluency by monolinguals and bilinguals. In addition, the production of the words over the 60 seconds of each trial was analyzed in terms of the number of words produced in 5-second time bins, following the procedure used by Rohrer, Wixted, Salmon, and Butters (1995). This method generates measures of intercept and slope to describe word production over the one minute time course. The intercept can be considered to index the initial linguistic resources available for the beginning of the trial and is largely determined by vocabulary knowledge. In contrast, the slope reflects how those resources are monitored and used over time and is largely determined by executive control; as the task progresses, more executive control is required to overcome the tendency to perseverate on previously generated responses and continue monitoring representational structures for more items. Thus, a flatter slope indicates that participants were able to maintain their performance across the response period despite greater interference towards the end of the trial, reflecting better executive control. Luo et al. found no slope differences between groups in the category condition
where little executive control is required, but higher intercepts\(^1\) for the monolinguals and the high proficiency bilinguals than for the low proficiency bilinguals. In the letter condition, Luo et al found dissociable effects of vocabulary knowledge and executive control. High proficiency bilinguals did not differ from monolinguals on the intercept and as in category fluency, both groups had significantly higher intercepts than the low proficiency bilinguals. This pattern is consistent with the interpretation that intercept reflects vocabulary resources. In contrast, both bilingual groups had a significantly flatter slope than the monolinguals irrespective of differences in intercept, thereby maintaining better performance across the timeframe despite growing interference as the task progresses.

To date, the investigation of executive control during the verbal fluency time course has been limited to young adults. Thus, it is unclear whether these findings generalize across the lifespan and when in development the benefits conferred by bilingualism in the letter fluency condition begin to emerge. It is also unknown whether the language group differences observed in young adulthood are maintained in older adults. The present study extends the investigation to include performance on verbal fluency tasks at four age groups: 7-year-olds, 10-year-olds, young adults and older adults. The young adults used in this analysis were the same monolinguals and high proficiency bilinguals reported in Luo et al. (2010). For the other three age groups, it was not possible to collect high and low language proficiency samples for each bilingual age group relative to their monolingual peers so the bilinguals in two of the new groups have vocabulary scores equivalent to those of their monolingual age mates and in one group have lower scores. Therefore, we present predictions for each of our new bilingual
samples relative to their monolingual control group based on the previously reported findings from young adults. There was no language group difference in English proficiency for the 7-year-olds, so if these young bilinguals can recruit control during a verbal fluency task, then the prediction is that they will perform better than monolinguals on letter fluency but similarly to monolinguals on category fluency. This advantage would be reflected in more items generated in the letter condition and a flatter slope in the letter time course. The 10-year-old bilinguals were less proficient in English than the monolinguals. Therefore, the prediction here is that the bilinguals will generate fewer category items, have a lower intercept on the category fluency and not differ on letter fluency similarly to the low proficiency bilinguals in Bialystok et al. (2008) and Luo et al. (2010). Finally, for the vocabulary-matched older adults, if the bilingual advantage is maintained in aging, the bilinguals will perform comparably to the monolinguals on the category fluency condition but better on the letter condition than the monolinguals.

Method

Participants

Data were compiled from different projects that used the verbal fluency task as part of a larger test battery. Participants were monolingual or bilingual and belonged to one of four age groups: 7-year-old children, 10-year-old children, younger adults or older adults (see Table 1 for sample sizes, mean ages and background information for each group). Monolingual participants spoke only English and had minimal or no knowledge of a second language. Bilinguals spoke English and another language fluently on a daily basis, and the adult samples reported English as their dominant language (see Table 2 for bilingual language profiles). Time course analyses from the young adults were previously
published in Luo et al. (2010) as the monolinguals and high proficiency bilinguals. Data from the low proficiency bilinguals in that study are not included in present analysis. The older adult sample was obtained from a study assessing aging, language experience and brain health. The monolingual children’s data were obtained from Bialystok, Peets, and Moreno (2014). The bilingual children were collected as new data for this study.

**Background Measures.** The Language and Social Background Questionnaire (LSBQ; Luk & Bialystok, 2013) was administered to all participants in the adult group and to parents of participating children in the two children groups to assess language experience. Since the data from the current study were drawn from independent projects, different measures were used to assess English vocabulary and non-verbal reasoning in each group. All children received the Peabody Picture Vocabulary Test third edition (Form B, PPVT-III; Dunn & Dunn, 1997) and the Raven’s Colored Progressive Matrices (Raven, Court, & Raven, 1990) which are both standardized around a mean of 100 and a standard deviation of 15. The young adults also completed the PPVT-III, but were given the Cattell Culture Fair Test (Cattell, 1957) to evaluate non-verbal reasoning. It is also standardized with a mean of 100 and a standard deviation of 15. The older adults were given the Shipley Vocabulary test (Shipley, 1940) as a measure of English receptive vocabulary. No non-verbal reasoning measure was administered but older adults were matched on Wisconsin Card Sorting Performance. All older adults were screened for cognitive impairment using an abbreviated version of the mini-mental state examination (MMSE), and all achieved at least 16 out of 17, indicating normal cognition.

*Verbal Fluency test (Delis-Kaplan Executive Function System, D-KEFS; Delis, Kaplan, & Kramer, 2001b).* Participants are asked to produce as many English words as
possible in 60 seconds. In the letter condition, there were three trials requiring responses to the letters F, A, and S. The adults were given four restrictions: (1) no variants of the same words, (2) no names of people, (3) no names of places, and (4) no numbers. No restrictions were given to children because it would make the task too difficult (see Kormi-Nouri et al., 2012 for similar procedures). Because the data were obtained from independent studies, the different semantic categories were used for the category condition across age groups. The children and the young adults were asked two categories, clothing items and girls’ names. The older adults provided responses to clothing items, animals and occupations. Only responses to the shared category (clothing items) were analyzed because the number of items generated is influenced by the category. For example, older adults produced more items in the animal category than in the other categories so the inclusion of this category would artificially inflate their scores relative to the younger adults. Additionally, although “girls’ names” is a category in the D-KEFS, it does not draw on vocabulary knowledge and in retrospect was a poor choice for this research. The only restriction on category fluency conditions was to say different words. Responses were recorded on a digital recorder. Number of correct items was obtained by subtracting incorrect responses (words that did not start with the specified letter or not in the designated categories) and repeated words from the total number of responses.

Data coding and analysis

Data coding was conducted following the procedures described in Luo et al. (2010). Trained research assistants processed the digital recordings of verbal fluency responses with Audacity® on a Windows platform. The research assistants first listened
to the recording to identify the correct responses and then recorded their associated timestamp. Based on the time-stamps, correct responses were grouped into 5-s bins over each 60-second trial, producing twelve bins. The following codes were used for each correct response in subsequent analyses: 1) serial number, indicating the serial position of a response in the trial; 2) subsequent-response latency measuring the time between the first response and the onset of each subsequent response, with the mean subsequent-response latency indicating the point at which half of the responses have been produced (also called “fulcrum”, Sandoval et al., 2010); and 3) bin number, indicating the 5-s bin into which the response falls. Subsequent-response latencies were calculated following the procedures described by Rohrer et al. (1995). Total number of correct responses and mean subsequent-response latency were obtained for each participant in each task. Longer mean subsequent-response latency indicates that performance extends later into the time course, but the interpretation of this variable depends on the total number of correct responses. If one group produces more correct responses than another group and has longer mean subsequent-response latency, then the interpretation is that the group has superior control (and equivalent or better vocabulary) and could continue generating responses longer. If one group produces fewer or equivalent correct responses but has longer mean subsequent-response latency, then the interpretation is that the control is more effortful because it took longer to generate the same or a fewer number of items. Therefore, both number of correct responses and mean subsequent-response latency need to be considered. Both correct responses and mean subsequent-retrieval latency were averaged across F, A, S trials in the letter condition, and taken from the clothing items trial for the category analysis.
Results

Background measures

Table 1 presents the background measures for each age and language group. The only significant difference between language groups was for the 10-year-olds in the vocabulary measure where the monolinguals scored higher than the bilinguals, $F(1, 43) = 8.75, p < .01$.

Number of Correct Responses

The mean number correct responses and mean subsequent latencies for each response are reported in Table 3. A 3-way analysis of variance was conducted on number of correct responses with age group (7-year-olds, 10-year-olds, young adults and older adults) and language group (monolingual and bilingual) as between-subject variables and condition (category and letter) as a within-subject variable. As expected, more items were generated in category fluency than letter fluency, $F(1, 157) = 112.39, p < .001$. There was a significant effect of age, $F(3, 157) = 71.57, p < .001$, in which both adult groups produced more items than the children, and the 10-year-olds generated more words than the 7-year-olds, with no difference between the young adults and the older adults. However, there was an interaction of age by condition, $F(3, 157) = 6.18, p < .01$, in which letter fluency performance improved into adulthood and remained stable in older age, but category fluency performance improved into adulthood and declined in older adults.

There was no main effect of language group for the number of items generated, $F < 1$, but there was a significant condition by language group interaction, $F(1, 157) = 6.92, p < .01$. Monolinguals and bilinguals did not differ on the category condition, $F(1,
157) = 1.24, n.s., but bilinguals produced more words than monolinguals on the letter condition, \( F(1, 157) = 3.69, p = .05 \). This interaction was further qualified by a three-way interaction of age, condition, and language group, \( F(3, 157) = 2.67, p = .05 \). Univariate analyses on the simple main effects were used to compare each language group within each fluency condition. In the letter condition young adult bilinguals outperformed monolinguals, \( F(1, 157) = 10.64, p < .001 \). As predicted, older bilinguals also produced more items in the letter condition than the monolinguals, \( F(1, 157) = 2.62, p = .05 \) (one-tailed test). In the category condition, 10-year-old bilinguals generated fewer words than monolinguals in the category condition, \( F(1,157) = 5.23, p < .05 \). There was a correlation between PPVT scores and number of items generated, \( r(1, 45) = .42, p < .01 \), in the 10-year-old sample supporting the interpretation that the observed performance differences were due to vocabulary scores. To investigate this possibility, a PPVT-matched sub-sample was analyzed (Monolinguals: \( N = 18, \text{PPVT} = 103.7, \text{SD} = 7.3 \); Bilinguals: \( N = 18, \text{PPVT} = 102.2, \text{SD} = 8.4 \)). In this subset, there was no longer any difference between the monolinguals (Category: \( M = 13.4, \text{SD} = 4.9 \); Letter: \( M = 9.3, \text{SD} = 2.8 \)) and bilinguals (Category: \( M = 11.3, \text{SD} = 4.3 \); Letter: \( M = 10.1, \text{SD} = 3.3 \)) on the category, \( F(1, 34) = 1.80, \text{n.s.} \), or the letter conditions, \( F < 1 \).

Consistent with the claim that letter fluency requires more executive control than category fluency, fewer words were produced on the letter task condition on the category condition. The magnitude of the difference between these conditions was calculated as a proportion of the category fluency score to reflect the additional resources needed for letter fluency production. Thus, a smaller proportion indicates better executive control, with the limiting case being equivalent performance on the two tasks. These proportion
scores are plotted in Figure 1. A 2-way ANOVA showed there was no main effect of age, \( F < 1 \), but there was a main effect of language group, \( F(3, 157) = 5.22, p < .03 \), and a marginal age by language group interaction, \( F(3,157) = 2.53, p = .06 \). The proportion difference was significantly smaller for bilinguals in both the 10-year-old, \( F(1, 157) = 3.81, p = .05 \), and the young adult groups, \( F(1, 157) = 8.65, p < .01 \). A one-tailed test also revealed a significant difference in proportion scores between the 10-year-old monolinguals (\( M = .22, SD = .3 \)) and bilinguals (\( M = .01, SD = .4 \)) in the PPVT-matched subset, \( F(1, 34) = 3.38, p = .04 \).

**Mean Subsequent Latencies**

A 3-way ANOVA for age, language group, and condition on mean subsequent-response latencies revealed a main effect of condition, \( F(1, 157) = 54.74, p < .001 \), in which longer latencies were observed on letter fluency than on category fluency, consistent with the more demanding nature of this condition. An effect of age, \( F(3, 157) = 3.74, p < .02 \), showed that the young adults had significantly longer mean subsequent latencies than did the other groups, but since they also produced more items, the interpretation is that they maintained performance further into the 60-second trial. There was a main effect of language group in which bilinguals exhibited longer latencies than monolinguals, \( F(1, 157) = 5.79, p < .02 \). Based on a significant age by group interaction, \( F(3, 157) = 2.54, p = .05 \), separate analyses were conducted on the adult and children data. Adults exhibited both main effects of age, \( F(1, 77) = 7.43, p < .01 \), and language group, \( F(1, 77) = 3.95, p = .05 \), wherein young adults and bilinguals produced longer mean subsequent latencies, but no age by language group interaction, \( F < 1 \). In contrast, in the children data, there was a significant interaction of language group by age, \( F(1, 79) \).
The 7-year-old bilinguals produced longer mean subsequent latencies than their monolingual peers, $F(1, 79) = 7.18, p < .01$. No differences were observed in the 10-year-olds, $F < 1$.

**Time Course Analysis**

Time courses analyses provide a more fine-grained depiction of how verbal responses are generated than do the single-point measure provided by mean subsequent latency. These analyses were conducted separately for the children and adults and for the letter and category fluency conditions because different patterns were expected for each. Each time course was fitted to multilevel models and the intercept and slope were analyzed as a function of language group and age group (7-year-olds vs. 10-year-olds in one analysis and young adults and older adults in another). The estimated functions from the multilevel models are presented in Table 4. Time courses are plotted in Figure 2 for category fluency and Figure 3 for letter fluency.

In the category fluency time course analysis for children, there was a significant effect of age on the intercept, $F(1, 80) = 21.48, p < .001$, and an age by group interaction, $F(1, 80) = 3.83, p = .05$. The effect indicates that the monolingual 10-year-olds had a higher intercept than the bilinguals of that age, $F(1, 43) = 4.20, p < .05$, consistent with their higher vocabulary scores. There was also a main effect of age on slope, $F(1, 920) = 21.79, p < .001$, in which 10-year-old children had a steeper slope than the 7-year-old children. For adults in the category fluency condition, young adults had significantly higher intercepts than the older adults, $F(1, 77) = 9.56, p < .01$. There were no differences between older and younger adults in the slope of the category fluency time course, and no language group differences in the intercept or the slope.
The time course analysis of letter fluency for children showed an effect of age for both intercept, $F(1, 80) = 31.48, p < .001$, and slope, $F(1, 920) = 13.19, p < .001$. Ten-year-old children had higher intercept and steeper slope than the 7-year-old children, but there were no differences between language groups on either measure. In the adult time course analysis for letter fluency, there were no age effects, but there was a significant effect of language group for both the intercept and slope. Bilinguals had a higher intercept, $F(1, 77) = 7.42, p < .01$, and flatter slope than monolinguals, $F(1,887) = 3.40, p = .06$. These effects did not interact with age. A summary of these effects is reported in Table 5.

**Discussion**

Language proficiency and executive control jointly contribute to successful lexical retrieval during language production, but their relative contributions were hypothesized to be differentially impacted by bilingualism. The goal of the present study was to explore how executive control involved in lexical retrieval develops and how bilingualism influences this development. To this end, developmental snapshots were presented by analyzing bilingual and monolingual performance on verbal fluency at four ages. Specifically of interest was determining the developmental point at which a bilingual advantage on letter fluency might emerge and the likelihood of its extension into older adulthood. It was predicted that bilinguals who had vocabulary equivalent to their monolingual peers would outperform monolinguals on letter fluency and that the two language groups would perform equivalently on category fluency. It was also predicted that bilinguals who had weaker vocabulary knowledge than their monolingual age-mates would produce fewer words than monolinguals in category fluency. The
results are largely consistent with these predictions.

To clarify the performance differences across the lifespan, the four age groups will be described separately and then the common patterns extracted to address more general issues. In the group of 7-year-olds, there was no difference between monolingual and bilingual children in their English standardized vocabulary scores. Therefore, it is not surprising that there were no language group differences on either the number of items generated or the time course analysis for the category fluency task. Since numerous studies have observed non-linguistic bilingual executive control advantages in young children (e.g., Bialystok, 2011; Carlson & Meltzoff, 2008; Poarch & van Hell, 2012; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011), it was more surprising that no bilingual advantage was observed on the letter fluency task. However, the letter condition relies on literacy, and these children were still early readers. Nonetheless, recall that Kormi-Nouri et al. (2012) found that Turkish-Persian bilinguals outperformed both Persian monolinguals and Kurdish-Persian bilinguals in Grade 1 on letter fluency but not in later grades. It is possible that any effects may be masked by variability in skills necessary to perform this difficulty task, skills that are automatic in adulthood. For example, at this age the variance associated with successful performance may involve individual differences in accessing lexical items and speech planning processes (Kavé, Kigel, & Kochva, 2008; Martins, Vieira, Loureiro, & Santos, 2007). Future research should investigate whether scaffolding in the form of strategy instruction may differentially help bilinguals. Hurks (2012) reported that monolingual children in Grade 3 did not benefit from instruction in strategy use on verbal fluency and suggested that the cognitive load was too large for them to implement the strategy. However, if bilingualism
supports greater readiness in the form of more efficient executive control, then instruction may provide the additional boost needed to produce group differences in the number of words produced.

Although no differences emerged on the number of responses, bilingual 7-year-olds did have longer mean subsequent latencies than their monolingual counterparts. Longer mean subsequent latencies in conjunction with equivalent or fewer items produced are consistent with the notion of word retrieval difficulties (Gollan et al., 2002; Luo et al., 2010). Since vocabulary scores did not differ between groups, it is unlikely that the difference emerged due to English proficiency differences. One possibility that has been raised is that the source of this increased difficulty in word production is cross-language interference that bilinguals experience through the concurrent activation of the two languages (see Portocarrero et al., 2007; Rosselli et al., 2000). Arguably, it is this need to recruit additional cognitive resources to resolve this conflict that makes word retrieval effortful but at the same time leads to executive control advantages.

Performance of the 10-year-old bilinguals most closely paralleled that of the low proficiency bilingual adults in Luo et al. (2010) who were also not matched with their monolingual peers on vocabulary level. As predicted, this vocabulary difference was only reflected in category fluency performance, so bilingual children generated fewer words and showed a lower intercept in the time-course analysis. Nonetheless, the proportion score between letter and category performance was smaller for bilingual 10-year-olds than monolingual 10-year-olds, suggesting that the bilingual children compensated for reduced linguistic resources, presumably through better executive control. Importantly, when the PPVT-matched subgroups were analyzed, group differences on the category
task no longer reached statistical significance and the proportion score difference between groups remained.

Consistent with predictions, no effect of bilingualism was observed on the category fluency task for the adults, but there was a significant effect of age; the older adults produced fewer words and had lower time course intercepts than the younger adults. This poorer category performance may be attributed to either verbal fluency performance declines associated with aging or with possible language proficiency differences between our older and younger adult groups. However, a dissociation between vocabulary knowledge and category fluency suggests that it is the former rather than the latter. Specifically, although the two age groups were given different vocabulary measures, the groups obtained similar standard scores. In order to achieve similar standard scores, norming procedures require older adults obtain higher raw scores. Thus, despite greater vocabulary knowledge, older adults had more difficulty efficiently accessing their lexical knowledge during verbal fluency than younger adults. This finding is in line with a number of previous studies finding an age-related decline in category fluency in monolinguals (e.g., Brickman et al, 2005; Parkin, Walter, & Hunkin, 1995; Tombaugh, Kozak & Rees, 1999). For example, Brickman and colleagues found greater age-related decline for category fluency than letter fluency. Taken together, these findings suggest that age and not bilingual status impacts category performance in adults.

In contrast to category fluency, adult letter fluency performance was related only to bilingual status and not to age, with a larger bilingual effect in the younger adults than in the older adult group. Recently Ljungberg et al. (2013) reported results from a large sample showing that older bilingual adults generated significantly more items during
letter fluency than older monolingual adults. In our study, effects were observed on the more sensitive time course measure; bilingual adults exhibited higher intercepts and flatter slopes than the monolinguals. This pattern was somewhat attenuated in the older adults, but it is worth noting on Figure 3 (panel B) that the older bilingual adults outperformed the young monolinguals. Bilingualism may boost performance because it brings executive control resources to the task, allowing adult bilinguals to outperform monolinguals, providing that vocabulary resources are equivalent.

Taken together, these results demonstrate that multiple factors impact verbal fluency performance including the nature of the task, age, language proficiency, and bilingual status. Each of these factors contributes to performance to varying degrees depending on the developmental time point under investigation. For example, the linguistic and cognitive resources available at 7-years-old may have been barely sufficient to perform the task so reliable differences in language status were not evident. Gains in vocabulary knowledge and executive control ability through childhood enable optimal performance in adulthood. Consistent with this, Gaillard et al. (2000) found that children activate similar brain regions as adults during verbal fluency tasks, but that this activation is greater and more diffuse in children, indicating that the brain networks underlying verbal fluency performance become more efficient during development. However, once adulthood is reached, age is less influential; verbal fluency performance is fairly stable through adulthood with only evidence of small declines in category fluency.

Bilingualism also impacts the developmental trajectory of verbal fluency performance. However, since only the letter task demands high levels of executive control, the effect of bilingualism was only observed in this condition. Beginning at 10-
years old, the executive control requirements for letter fluency were more easily handled by bilinguals than monolinguals, with a bilingual advantage on this task emerging in adulthood. This dissociation is most clearly demonstrated by the age effect found in the adult category fluency time course and the bilingual status effect found in the adult letter fluency time course. Thus, category fluency is primarily impacted by age and vocabulary knowledge whereas letter fluency is influenced by vocabulary knowledge and executive control, and as a consequence, bilingualism.

Differences in how bilinguals engage executive control during verbal fluency have implications for everyday bilingual language production. Arguably, since language production requires both language knowledge (e.g., vocabulary & syntax) and the engagement of executive control processes to manipulate these representations, bilingual executive control ability may compensate for poorer language proficiency. Work by Peets and Bialystok (in press) highlights the dissociation between formal measures of language proficiency and conversational measures. Peets and Bialystok found that although bilingual kindergarten students were less proficient in English than their monolingual counterparts according to standardized measures, analyses of bilinguals’ academic narratives revealed no differences in their vocabulary and grammar use. Additionally, there was no group difference in the sophistication of their discourse (e.g., organization, ideas, use of language). Essentially, the bilingual children were able to achieve the same performance level with less formal language knowledge, suggesting that bilinguals may be able to deploy their resources more effectually during everyday language production.

In conclusion, the present study is the first to report verbal fluency performance at different ages across the lifespan for both monolinguals and bilinguals. By presenting
snapshots of these two language groups at different ages we can begin to understand both how executive control develops and the role of language proficiency in lexical retrieval. In general, the results are consistent with the interpretation that vocabulary proficiency impacts both category and letter fluency. However, executive control impacts letter fluency more than category fluency and is sensitive to the influence of bilingualism. Thus, the relation between monolinguals and bilinguals on these neuropsychological tests depend crucially on levels of vocabulary knowledge in the language of testing. Unfortunately, when comparing the two groups in most studies, formal assessments of vocabulary level are rarely undertaken. More importantly, however, verbal fluency tests are used as standard neuropsychological assessment tools in clinical settings. Understanding that the underlying demands of these tests for vocabulary knowledge and executive control are differentially available in individuals with different experiences such as bilingualism is crucial information and should be incorporated into standard interpretations of test results.
Notes

1Luo et al. (2010) analyzed both intercept and initiation parameter. They present the data for initiation parameter but note that the results of the analyses were the same for both variables.
References


Table 1

Background Measures by Age and Language Group (Standard Deviations in Brackets).

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>English</th>
<th>Non-Verbal IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Vocabulary</td>
<td></td>
</tr>
<tr>
<td><strong>7-year-olds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolingual</td>
<td>16</td>
<td>7.7 (0.3)</td>
<td>108.7 (12.8)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bilingual</td>
<td>23</td>
<td>7.7 (0.4)</td>
<td>105.6 (11.6)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>10-year-olds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolingual</td>
<td>22</td>
<td>10.6 (0.5)</td>
<td>107.4 (10.5)&lt;sup&gt;a,c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bilingual</td>
<td>23</td>
<td>10.6 (0.4)</td>
<td>97.0 (12.7)&lt;sup&gt;b,c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Younger Adults</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolingual</td>
<td>20</td>
<td>20.7 (1.3)</td>
<td>106.4 (4.0)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bilingual</td>
<td>20</td>
<td>21.1 (1.3)</td>
<td>107.5 (3.5)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Older Adults</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolingual</td>
<td>20</td>
<td>70.9 (2.6)</td>
<td>107.8 (6.1)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bilingual</td>
<td>21</td>
<td>71.1 (3.8)</td>
<td>104.3 (11.3)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Peabody Picture Vocabulary Test-III  
<sup>b</sup>Shipley I (Normed based on the Shipley-II)  
<sup>c</sup>Raven’s Colored Progressive Matrices  
<sup>d</sup>Cattell Culture Fair Test  
<sup>e</sup>Monolinguals > Bilinguals
Table 2

*Bilingual Language Profile by Age Group*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>% English as L1</th>
<th>% Non-English use at home&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Age of L2 Acquisition</th>
<th>English Speaking Ratings&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Non-English Speaking Ratings&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-year-olds</td>
<td>32 daily</td>
<td>3.2 (1.5)</td>
<td>-</td>
<td>3.5 (1.0)</td>
<td></td>
</tr>
<tr>
<td>10-year-olds</td>
<td>30 daily</td>
<td>3.8 (2.3)</td>
<td>-</td>
<td>2.7 (1.1)</td>
<td></td>
</tr>
<tr>
<td>Younger Adults</td>
<td>100 17.7 (21.3)</td>
<td>2.9 (4.2)</td>
<td>9.3 (1.8)</td>
<td>6.3 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Older Adults</td>
<td>30 32.2 (33.9)</td>
<td>8.8 (5.9)</td>
<td>9.6 (0.7)</td>
<td>8.5 (2.2)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>In the Children's questionnaire, parents were given a scale to rate their child's non-English language use (daily, weekly, monthly, occasionally, other or NA). Adults rated the percentage of their Non-English language use on a daily basis.

<sup>b</sup>Adults were asked to rate their English ability on a scale from 0 to 10 where 10 was native-like ability. Not all parents were in a position to rate their child’s English ability and this question was not asked.

<sup>c</sup>The Non-English Speaking rating was out of 5 for children and out of 10 for adults.
Table 3

*Mean Scores and Standard Deviations for Mean Number of Correct Responses and Mean Subsequent Response Latencies (in seconds)*

<table>
<thead>
<tr>
<th>Task</th>
<th>Number Correct Responses</th>
<th>Mean Subsequent-Response Latencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monolingual</td>
<td>Bilingual</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>8.5 (2.9)</td>
<td>8.4 (3.0)</td>
</tr>
<tr>
<td>Letter</td>
<td>6.8 (1.9)</td>
<td>6.0 (2.2)</td>
</tr>
<tr>
<td>10-year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>13.6 (4.9)</td>
<td>10.8 (4.3)</td>
</tr>
<tr>
<td>Letter</td>
<td>9.8 (3.6)</td>
<td>9.4 (3.2)</td>
</tr>
<tr>
<td>Younger Adults</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>19.9 (4.6)</td>
<td>19.0 (5.7)</td>
</tr>
<tr>
<td>Letter</td>
<td>11.8 (3.5)</td>
<td>15.3 (4.0)</td>
</tr>
<tr>
<td>Older Adults</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>17.0 (3.0)</td>
<td>17.9 (3.5)</td>
</tr>
<tr>
<td>Letter</td>
<td>12.9 (3.5)</td>
<td>14.7 (4.6)</td>
</tr>
</tbody>
</table>
Table 4

*Best Fitting Multilevel Model Functions for the Time Course of Verbal Fluency Output*

<table>
<thead>
<tr>
<th>Task</th>
<th>Category</th>
<th>Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Children</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-year-old monolingual</td>
<td>$y = 2.03 - 0.77\ln(t)$</td>
<td>$y = 1.32 - 0.43\ln(t)$</td>
</tr>
<tr>
<td>7-year-old bilingual</td>
<td>$y = 1.83 - 0.62\ln(t)$</td>
<td>$y = 1.31 - 0.43\ln(t)$</td>
</tr>
<tr>
<td>10-year-old monolingual</td>
<td>$y = 2.89 - 1.01\ln(t)$</td>
<td>$y = 1.94 - 0.62\ln(t)$</td>
</tr>
<tr>
<td>10-year-old bilingual</td>
<td>$y = 2.80 - 1.09\ln(t)$</td>
<td>$y = 1.76 - 0.54\ln(t)$</td>
</tr>
<tr>
<td><strong>Adults</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger monolingual</td>
<td>$y = 3.64 - 1.14\ln(t)$</td>
<td>$y = 2.29 - 0.73\ln(t)$</td>
</tr>
<tr>
<td>Younger bilingual</td>
<td>$y = 3.51 - 1.11\ln(t)$</td>
<td>$y = 2.36 - 0.60\ln(t)$</td>
</tr>
<tr>
<td>Older monolingual</td>
<td>$y = 2.96 - 0.92\ln(t)$</td>
<td>$y = 2.18 - 0.66\ln(t)$</td>
</tr>
<tr>
<td>Older bilingual</td>
<td>$y = 3.27 - 1.07\ln(t)$</td>
<td>$y = 2.27 - 0.63\ln(t)$</td>
</tr>
</tbody>
</table>

Multilevel Model: logarithmic function estimates obtained from multilevel modeling with all observations.
Table 5

*Summary of Language Group Effects*

<table>
<thead>
<tr>
<th>Category</th>
<th>7-year-olds</th>
<th>10-year-olds</th>
<th>Younger</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Responses</strong></td>
<td>M = B</td>
<td><strong>M &gt; B</strong></td>
<td>M = B</td>
<td></td>
</tr>
<tr>
<td><strong>Mean Subsequent Latencies</strong></td>
<td><strong>M &lt; B</strong></td>
<td>M = B</td>
<td><strong>M &lt; B</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Time Course Intercept</strong></td>
<td>M = B</td>
<td><strong>M &gt; B</strong></td>
<td>M = B</td>
<td></td>
</tr>
<tr>
<td><strong>Time Course Slope</strong></td>
<td>M = B</td>
<td>M = B</td>
<td>M = B</td>
<td></td>
</tr>
</tbody>
</table>

**Letter**

<table>
<thead>
<tr>
<th>Category</th>
<th>7-year-olds</th>
<th>10-year-olds</th>
<th>Younger</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Responses</strong></td>
<td>M = B</td>
<td><strong>M &lt; B</strong></td>
<td>M &lt; B</td>
<td><strong>M &lt; B</strong></td>
</tr>
<tr>
<td><strong>Mean Subsequent Latencies</strong></td>
<td><strong>M &lt; B</strong></td>
<td>M = B</td>
<td><strong>M &lt; B</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Time Course Intercept</strong></td>
<td>M = B</td>
<td>M = B</td>
<td>M &lt; B</td>
<td></td>
</tr>
<tr>
<td><strong>Time Course Slope</strong></td>
<td>M = B</td>
<td>M = B</td>
<td><strong>M &gt; B</strong></td>
<td></td>
</tr>
</tbody>
</table>

Note. Effects that are positioned directly below a group apply to that group only and effects that are centered apply to the higher order grouping.

*a* marginal effect
Figure Captions

*Figure 1.* Mean proportion difference scores between words generated in category fluency and letter fluency and standard error by age group and language group.

*Figure 2.* Number of items produced as a function of time in the category task for A children and B adults. Best fit lines are logarithmic functions.

*Figure 3.* Number of items produced as a function of time in the letter task for A children and B adults. Best fit lines are logarithmic functions.
Figure 1.
Figure 2

A) Children

B) Adults
Figure 3

A) Children

B) Adults