Estimating Vocabulary Size with Verbal Fluency Scores

Daphnée Dubé

School of Communication Sciences and Disorders

McGill University, Montreal

April 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Abstract

Speech-language pathologists often see clients who are speakers of one or more languages that they cannot assess for different reasons. We proposed using a well-known experimental task, verbal fluency, typically used in research as a measure of executive function, to estimate vocabulary knowledge when normed assessments are not possible. In a series of analyses, 193 monolingual and multilingual children and adolescents completed tests of vocabulary knowledge and of verbal fluency. Participants were exposed to different language combinations and tested in Canada (English, French) and Iceland (Icelandic, English, Other languages). First, we examined the ability of verbal fluency measures to predict vocabulary size in three languages in which vocabulary size could be measured directly (n = 193). Predictors included well-established measures (total correct words, mean cluster size, number of switches) and a new, modified measure of total correct words on two types of verbal fluency tasks (semantic and phonemic). Using simple linear regressions on the data of participants who were administered both the verbal fluency task and vocabulary tests, both measures of total correct words (traditional and modified) were the best predictors of vocabulary size. Innovatively, we then extended the analysis to languages in which vocabulary measurement was not possible, in which case verbal fluency could be most useful clinically (n = 36). In those languages, self-ratings of performance were used instead of vocabulary tests. Again, verbal fluency strongly predicted performance. In addition to examining the value of traditional measures of verbal fluency, we examined responses in novel ways, including the use of infrequent words, production of translation equivalents, and longitudinal performance and their implications for vocabulary knowledge. This thesis supports the use of the verbal fluency task as a quick and simple tool for speech-language pathologists to use as a rough estimate of vocabulary knowledge.

Résumé

Les orthophonistes travaillent souvent avec des clients chez qui l'une ou plusieurs de leurs langues ne peuvent être évaluées pour de différentes raisons. Nous avons proposé l'usage d'une tâche expérimentale bien établie, la tâche de fluence verbale, à utiliser lorsque l'utilisation d'un outil normalisé n'est pas possible. À travers d'une série d'analyses, 193 enfants et adolescents unilingues et multilingues ont été évalués avec des évaluations de vocabulaire et de fluence verbale. Les participants étaient exposés à de différentes combinaisons de langues et ont été testés au Canada (anglais, français) et en Islande (islandais, anglais, autres langues). D'abord, des mesures de fluence verbale ont été évaluées en tant que prédicteurs de taille de vocabulaire dans des langues pour lesquelles la taille de vocabulaire pouvait être mesurée directement (*n* = 193). Les prédicteurs ont inclus des mesures bien établies (nombre de mots corrects, taille moyenne de *cluster*, nombre de *cluster switches*) et une nouvelle mesure modifiée de nombre de mots corrects sur deux tâches de fluence verbale (sémantique et phonémique). En utilisant des régressions linéaires simples sur les données des participants qui ont complété les tâches de fluence verbale ainsi que les évaluations de vocabulaire, les deux mesures de mots corrects (traditionnelle et modifiée) étaient les meilleurs prédicteurs de scores de vocabulaire. De manière innovante, nous avons ensuite étendu nos analyses à des langues pour lesquelles la taille de vocabulaire ne pouvait être mesurée (*n* = 36). Dans ces langues, des auto-évaluations de la performance ont été utilisée au lieu d'évaluations de vocabulaire. Encore une fois, la fluence verbale a prédit la performance. En plus d'évaluer la valeur de mesures traditionnelles, nous avons étudié les réponses de nouvelles manières, incluant l'usage de mots non-fréquents et de traductions et la performance longitudinale, ainsi que l'implication de ces mesures pour les connaissances de vocabulaire. Cette thèse supporte l’usage des tâches de fluence verbale comme outil rapide et simple à utiliser comme estimation de connaissances de vocabulaire.

Acknowledgements

I would like to thank my supervisor, Dr. Elin Thordardottir, for her guidance and support throughout this thesis year so unlike any other, and for helping feed my passion for research. Thank you to Dr. Aparna Nadig for her feedback and suggestions.

I would also like to thank the members of the Child Language Development and Disorders Lab and the Rannsóknarstofa í málþroska og fjöltyngi at the Academy of Reykjavik who participated in recruiting and testing participants. Special thanks to Ludivine Plez, Inga Ragnarsdóttir, and Heather Ragnars for hours of transcribing and reliability coding and to Myrto Brandeker and Birkir Már Viðarsson for their help throughout and for responding to my many emails. I am grateful to the participants without whom this project would not have been possible.

Thank you to Nick, my partner, who supported me and was the best colleague I could have asked for during this confinement-thesis year. Thank you to my parents, for their belief in me since the day I said I wanted to complete my dream program. To my friends and family, I appreciate your patience and look forward to spending more time with you.

This study was funded by master's bursaries awarded to the author by the McGill Faculty of Medicine Research Bursary Program and the Centre for Research in Rehabilitation of Greater Montreal (CRIR). The thesis used data collected in studies funded by grants from SSHRC (The Social Sciences and Humanities Research Council of Canada) and RANNÍS (The Icelandic Centre for Research) awarded to Elin Thordardottir.

Author Contributions

Author contributions included the following:

- Testing a subset of participants during a McGill Medicine Summer Bursary term,
- Transcribing and scoring English and French trials for all newly recruited participants (including English trials for Icelandic participants),
- Re-scoring a subset of data for previously recruited participants,
- Scoring Icelandic data from translated transcriptions,
- Training reliability scorers and calculating reliability scores,
- Organizing data and running analyses,
- Writing the thesis.

Table of Contents

List of Tables and Figures

Introduction

The field of speech-language pathology is relatively young and has experienced much growth in the past few decades. Many theoretical and practical advances have been achieved, predominantly in English and with a focus on children, but significant research has also been conducted in other languages, such as French, Spanish, Mandarin, and Icelandic, among others. Such advancements include the creation of standardized tests to assess the different components of language. These standardized tests allow for a common procedure to be followed by examiners, as well as for comparison to a group norm used to help identify or rule out developmental language disorders and to estimate the client's level of knowledge.

One area for which standardized tests have been created is the assessment of vocabulary skills, which can be assessed receptively (identifying an item from a set) and expressively (labeling items). In Quebec, Canada, speech-language pathology clients are often assessed in English or French. Receptive vocabulary can be assessed in English, for example, with the Peabody Picture Vocabulary Test – Third Edition (PPVT-III; Dunn & Dunn, 1997), or by later versions of the PPVT, and in French with its French-language equivalent, the *Échelle de vocabulaire en images Peabody* (EVIP, Dunn et al., 1993). These tests require the examinee to correctly identify an item or action from a set of 4 images. Expressive vocabulary can be assessed in English with the Expressive Vocabulary Test – Second Edition (EVT; Williams, 1997) and the Expressive Vocabulary subtest from the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4; Semel, Wiig & Secord, 2003), among others. French expressive vocabulary can be assessed with the Expressive Vocabulary subtest from the *Évaluation clinique des notions langagières fondamentales* in French (CELF-CDN-F; Semel et al., 2009), but only until age 9, after which point there exists no standardized vocabulary test normed for Quebec populations. Tests of expressive vocabulary frequently require examinees to correctly label pictured items or actions.

Quebec and Canada are also the homes of many speakers of indigenous languages and of immigrants who speak a language other than English or French. About 21% of the population of Canada learned a language other than English or French as their first language, and an estimated 27% have knowledge of languages other than English or French (Statistics Canada, 2016). For many of these languages, standardized tests do not exist. Although clients and clinicians may speak a common language (often English or French), it is important to gather information in all of the client's languages in order to get an accurate portrait of their linguistic abilities. When no standardized test exists in a given language, some clinicians might opt for non-normed translated versions of tests or 'home-made' tests if one is available or if the clinician speaks this language. Another option includes using naturalistic observation of the client. If the clinician does not speak this language, they might choose to use adapted materials and work with an interpreter. These solutions come with their own obstacles, the main one being that these tests are typically not standardized and have not been normed, thus making it more difficult to draw accurate conclusions. Given the lack of vocabulary tests for numerous populations, this skill cannot be formally assessed in many individuals. Should an appropriate vocabulary test be available for use in multiple languages, this issue could be alleviated.

We propose trying to contribute to solving this issue by examining a way to estimate vocabulary knowledge using, as our tool, an established experimental task: the verbal fluency task. This task involves producing words that conform to a specific criterion. This criterion can be based on a meaning category (called semantic or category fluency) or a word-initial sound or letter (called phonemic/phonological or letter fluency), within a specified time limit, often 60 seconds. Verbal fluency tasks therefore demand specific skills and knowledge to be used by the participant. They require individuals to attend to and select items conforming to the appropriate semantic or phonemic set, while inhibiting items that do not match the criterion, thus relying on executive functions. To date, verbal fluency has been used primarily to assess executive functioning and compare executive control in different populations (e.g., Brandeker & Elin Thordardottir, under review; Friesen et al., 2015; Hurks et al., 2010; Kormi-Nouri et al., 2012; Troyer, 2000). Clinically, verbal fluency has also shown value as an indicator for neurological disorders in adults in multiple meta-analyses (e.g., Bokat & Goldberg, 2003; Henry & Crawford, 2004a; Henry et al., 2004). Despite its frequent use in studies of executive functioning, verbal fluency also has important links to another skill: vocabulary knowledge. Verbal fluency tasks, by their very nature, require the participant to produce appropriate lexical items. In the above-mentioned clinical studies, although a link between verbal fluency and language was noted, the focus was not on predicting the language abilities of these patients so much as it was to indicate that neurological deficits were present (e.g., Bokat & Goldberg, 2003; Henry & Crawford, 2004a; Henry et al., 2004). Verbal fluency's value as a clinical estimation of vocabulary knowledge has not yet been explored. Should verbal fluency measures successfully estimate vocabulary size, this task could become a valuable assessment tool for this purpose.

In this thesis, the use of verbal fluency tasks as clinical indicators of vocabulary knowledge was assessed in a series of analyses (Analyses 1 to 6 below) including data from three languages (English, French, and Icelandic) and subsequently by extending findings to a pool of other languages. Participants in this series of analyses were drawn from two ongoing large-scale studies conducted in Montreal, Canada and in Reykjavik, Iceland. These two testing locations allowed the study of verbal fluency in two significantly different language contexts and across different language combinations.

Literature Review

Verbal fluency tasks have been used in many studies and are quite standardized in regards to how the tasks are administered (provide a criterion, allow a set amount of time, usually have participants speak words out loud) and how they are scored. The most common measure obtained from these tasks is the total number of correct words produced. In order to be considered a correct word, it needs to, evidently, adhere to the provided criterion. For example, naming a fruit on a task of animal fluency would be scored as incorrect. However, additional scoring criteria have also been laid out and frequently used (e.g., Troyer et al., 1997). For example, on semantic fluency tasks (often animal fluency), if a participant provides both a superordinate category item and an exemplar (e.g. bird and robin), a point is only awarded for the exemplar. If sex (hen and rooster) or age (dog and puppy) variants are provided, only one of these items is awarded a point. For phonemic fluency tasks, proper names and repeated root words with different endings are traditionally not awarded any points. These guidelines are typically followed in studies of verbal fluency.

As mentioned, verbal fluency has frequently been used for the purpose of assessing executive functioning. The construct validity of verbal fluency tasks has been studied through factor analyses, in attempts to determine whether verbal fluency is actually more a measure of language versus of executive functioning. In one study, this was assessed in healthy college-age participants, with results showing both types of verbal fluency (semantic and phonemic) having links to both executive functioning and word knowledge (Aita et al., 2018). In another study, this question was assessed in adults with neurological problems, psychiatric problems, or no diagnosis (Whiteside et al., 2016). Measures of both semantic and phonemic verbal fluency loaded only on the language factor (labeled as such given significant loadings on naming and vocabulary assessments), both when patients with different diagnoses were included in analyses together or separately. These findings support the idea that vocabulary knowledge should be considered in verbal fluency tasks. Taken together, both executive functioning and vocabulary knowledge appear to play a role in verbal fluency performance.

An area of research which has frequently used verbal fluency tasks is language processing in bilinguals, who are reported to differentially compare to monolinguals on vocabulary knowledge and executive functioning, the two skills that verbal fluency tasks recruit. On one hand, bilinguals have been found to score lower on tests that more closely target vocabulary knowledge, which has been attributed to smaller vocabulary sizes in each language (e.g., Oller & Eilers, 2002; Pearson et al., 1993) and to interference from the other language during lexical retrieval (as discussed by Bialystok, 2009). On the other hand, bilinguals have been reported to perform better than monolinguals on tests of executive functioning, which has been linked to bilinguals' constant need to attend to words in the appropriate language, while inhibiting words from the other language (e.g., Bialystok, 2009). Verbal fluency tasks therefore allow researchers to compare executive functioning between monolinguals and bilinguals, who may differ in these abilities.

Types of Verbal Fluency

As mentioned, different types of verbal fluency exist: semantic and phonemic. Both are relevant to consider in this search for an estimate of vocabulary knowledge, and there is an interest in determining whether one of these might be particularly useful for our purposes. Insight into this question might be obtained from comparisons between the performance of monolinguals and bilinguals on these tasks. Bilinguals have typically been found to perform similarly to or outperform monolinguals on tests of phonemic fluency but to perform worse than monolinguals on tests of semantic fluency. This has been observed in younger and older adults (e.g., Friesen et al., 2015; see Bialystok, 2009) and in children (Brandeker & Elin Thordardottir, under review; Friesen et al., 2015; Kormi-Nouri et al., 2012). Based on such results, it has traditionally been proposed that phonemic fluency is more closely related to executive function abilities, an area of relative advantage for bilinguals, whereas semantic fluency is more closely related to language-specific vocabulary knowledge, where bilinguals are reported to be at a disadvantage (e.g., Bialystok, 2009). However, one issue that needs consideration is the possibility that group differences in verbal fluency might have been driven by differences in vocabulary size. In order to specifically target executive functioning, which has typically been the purpose of these studies, differences in vocabulary scores need to be controlled for – something that many studies have not done.

Studies are available that have controlled the effect of vocabulary size by recruiting bilinguals who have the same vocabulary size as monolinguals in the language in which verbal fluency was tested. These studies have found that group differences showing monolinguals outperforming bilinguals on semantic fluency tasks were no longer observed; however the bilingual advantage on phonemic tasks remained (Friesen, et al., 2015; Paap et al., 2017) or bilinguals outperformed monolinguals on tasks of both semantic and phonemic fluency (Escobar et al., 2018). The former was also true when bilinguals were also matched to the monolingual comparison group on age, education, and non-verbal intelligence (Patra et al., 2020). The common finding was that monolinguals no longer outperformed bilinguals on semantic fluency tasks. These types of findings have led researchers to conclude that vocabulary knowledge plays a role in verbal fluency performance, such that it obscures the measurement of executive functioning and must therefore be controlled. In contrast to these studies in which the main interest was in executive function, the goal in this thesis is rather to zone in on vocabulary knowledge to examine whether its close links to semantic and phonemic fluency can be useful for clinical purposes.

Types of Verbal Fluency and Vocabulary Size

Despite the general focus on executive function, some studies of verbal fluency have investigated, beyond group comparisons, the relationships between verbal fluency measures and vocabulary size, particularly in terms of differences between semantic and phonemic fluency. Findings have been mixed, for different possible reasons. Vocabulary scores have been found to be correlated with the number of words produced on tasks of semantic but not on tasks of phonemic fluency in monolingual and bilingual children and in monolingual adults (Friesen et al., 2015; Shao et al., 2014). In another study of monolingual adults, vocabulary scores did correlate with the number of words produced on phonemic fluency tasks, but that study used different vocabulary measures than those typically used (synonym and antonym tests and selecting items with similar meanings) and required participants to write rather than speak their responses (Hedden et al., 2005), which may have tapped into different skills than in the other studies. However, this link with phonemic fluency is also supported by a study with French-English bilingual children, in which language-specific correlations were completed, showing French and English vocabulary size correlating with French and English verbal fluency performance respectively, both for semantic and phonemic fluency (Brandeker & Elin Thordardottir, under review). The relation between vocabulary scores and verbal fluency measures has also been explored by assessing vocabulary scores as a predictor of verbal fluency scores. In bilingual children, receptive vocabulary scores made significant predictions of both semantic and phonemic fluency scores (Escobar et al., 2018). However, in monolingual adults, vocabulary scores predicted scores on neither type of verbal fluency task (Shao et al., 2014). This might be explained by a potential ceiling effect in verbal fluency that might occur in adults but not children. Further, in bilingual children, cumulative exposure to each language did not predict verbal fluency performance, which suggests that task performance is particularly closely linked to the vocabulary knowledge someone has, rather than simply how much exposure to a language someone has received (Brandeker & Elin Thordardottir, under review). Importantly these studies used vocabulary scores to predict verbal fluency performance, but not the other way around. Despite the inherent need for vocabulary knowledge to perform these tasks, the link between vocabulary and verbal fluency has not received sufficient investigation, and this investigation should include both semantic and phonemic fluency measures as contenders for useful estimators of vocabulary size.

Additional Verbal Fluency Measures

In response to the finding by Whiteside et al. (2016) of the important link between both types of verbal fluency and language, Patra et al. (2020) argued that, beyond controlling for certain variables, the reason for such a finding is the inclusion of only one measure of verbal fluency: total number of correct words produced. Patra et al. further proposed that multiple measures of verbal fluency should be considered for a more comprehensive understanding of performance on these tasks. While the majority of studies only assess the total number of correct words produced, some studies have addressed the use of search strategies during verbal fluency tasks, including forming clusters and switching between clusters (e.g., Patra et al., 2020; Troyer et al., 1997). Clusters are groups of words produced consecutively that are related in some way (e.g., ‘cat, tiger, lion’ form a cluster because they are all felines). Switching between clusters refers to changing from listing words in a cluster to producing an unrelated word (e.g., ‘cat, tiger, lion, bear’: the switch occurs after ‘lion’ because a bear is not a feline). In monolingual children, it has been argued that mean cluster size is more closely related to vocabulary knowledge, whereas switching is more closely related to executive function and has a greater effect on performance on verbal fluency tasks (Filippetti & Allegri, 2011). In adults, no differences between monolinguals and bilinguals were found for switching in general nor for cluster size on the semantic fluency task, but bilinguals produced larger clusters than monolinguals on phonemic fluency tasks, which led to more correct words (Patra et al., 2020). It was suggested that forming large clusters may be a strategy used by bilinguals in a more demanding task. Very few studies exist assessing these measures in bilingual children.

A previous study in our lab examined this in a sample of 88 typically developing monolingual and bilingual first and third graders who varied widely in their previous relative exposure to French and English (Brandeker & Elin Thordardottir, under review). On the group level, for the measure of number of switches between clusters, no differences between monolingual and bilingual children were observed. For mean cluster size, no differences were found for phonemic fluency, but on the semantic fluency task, monolinguals formed larger clusters than bilinguals. This finding parallels the study’s results for the measure of total words produced. Regression analysis for the entire group revealed that both cluster size and number of switches were strong predictors of the number of words produced. Previously, it had been suggested, based on monolingual children, that cluster size is a measure of vocabulary knowledge (Filippetti & Allegri, 2011). In Brandeker and Elin Thordardottir, a regression analysis of the bilingual children only compared to the entire group suggested that the bilingual children achieved their total vocabulary scores by relying on switches rather than large cluster sizes. While the results for the entire group show that both strategies are viable, the bilingual children had access primarily to the switching strategy (however not more than the monolingual children). Given that they also had lower vocabulary sizes than the monolingual group, this can support the previous claim by Filippetti and Allegri that cluster size is associated with vocabulary knowledge. The same question was investigated in a smaller sample of 26 Autistic monolingual and bilingual 5- to 10-year-olds, compared to 26 neurotypical peers, of particular interest to the current study. No differences were found between neurotypical monolinguals and bilinguals on any of these measures (Gonzalez-Barrero & Nadig, 2016). Findings are therefore inconsistent.

Other measures have been used in the literature in studies with adults. These include time-course analysis (Luo et al., 2010), as well as initial words produced and difference scores between semantic and phonemic fluency (Patra et al., 2020). However, these go beyond the scope of this thesis.

Effect of Dominance

During a language assessment in bilinguals, other than obtaining vocabulary measures, it is also valuable to determine which language is their dominant language. Dominance is important clinically because it indicates which language is being used more often, which can make it the more important target of intervention. When thorough measurement of both languages is not possible, it is also very valuable to be able to establish which language is the stronger one – this helps interpret the test results that can be obtained. Some parents may have an idea of their child's dominant language, which can be in line with their relative exposure to each language, but this may be unclear given uncertainties about language use in childcare settings or in school. It would therefore be helpful to have a way to help determine language dominance, possibly by using verbal fluency tasks. Comparisons between bilingual children's performance on verbal fluency tasks in their two languages have been only infrequently made. In their study of 1st and 3rd graders tested in both English and French, Brandeker and Elin Thordardottir (under review) reported that total correct words produced showed no differences between languages in either the first grade or third grade. However, performance in each language was compared by including all bilinguals together without considering specific dominance groups (English-dominant, French-dominant), which may have cancelled out potential effects because the bilingual group included children with varying relative performance in French and English. Additional comparisons are warranted to determine whether comparing verbal fluency performance in the two languages of a bilingual child could detect existing differences in dominance between the languages. Documenting linguistic experiences including language dominance is difficult in certain contexts. In Quebec, some children have a definite language dominance, whereas others do not and might be similarly proficient in each of their languages. The same has been observed in Iceland, where people are L1 or L2 speakers of Icelandic, most gain incidental exposure of English through the media, and many have another home language (Elin Thordardottir, 2021). It is of interest whether performance on verbal fluency tasks can detect language dominance in different linguistic contexts.

Preliminary Findings

In our previous pilot study, French-English bilingual adolescents completed verbal fluency and vocabulary tests in both of their languages (Dubé & Elin Thordardottir, 2019). Verbal fluency scores were used to estimate vocabulary scores, using simple linear regressions. Results were promising enough, with significant language-specific regressions being observed, that we decided to examine these questions on a larger scale and to include additional measures of verbal fluency in the current thesis.

Gaps in the Literature and Novel Measures

Research has clearly shown links between verbal fluency and vocabulary knowledge (e.g., Brandeker & Elin Thordardottir, under review; Friesen et al., 2015; Kormi-Nouri et al., 2012). However, measures derived from verbal fluency tasks have not yet been used to estimate vocabulary size. This was attempted in the current study. Studies have also typically only assessed participants in one language (but see Brandeker & Elin Thordardottir, under review). Beyond attempting to make estimates in the same language, cross-language comparisons (e.g., English verbal fluency to estimate French vocabulary scores) are important in order to confirm whether or not estimations are language-specific. If they are not, then they may be related to a more general language ability rather than vocabulary knowledge in the language of the verbal fluency test, or to some other factor. If this is the case, the clinical utility of verbal fluency to estimate vocabulary size is limited to an overall vocabulary estimate. For this reason, the ability of verbal fluency to predict vocabulary size was examined in both languages of bilingual participants in this study. It is possible that typological differences between languages, such as the greater ease of forming compound words in English than in French, may influence the degree to which vocabulary size predicts verbal fluency, thus possibly detracting from the usefulness of verbal fluency for this purpose, particularly for phonemic fluency, which was kept in mind (e.g., participants can say “sun, sunscreen, sunlight, sunglasses”). However, this was not predicted to be an issue for semantic fluency measures, since this task requires naming specific category items (e.g., cat, dog, bird) which cannot be built on in the same way that words following the same sound-based criterion can.

In addition, previous studies have focused on languages for which vocabulary knowledge could be measured. However, speech-language pathologists frequently assess bilinguals whose second or often first language they do not know and/or for which formal tests are lacking, resulting in an incomplete assessment. Should the relationship between verbal fluency and vocabulary size be found to be language-specific for those languages in which both verbal fluency and vocabulary score are available, this would establish the validity of using this method to estimate vocabulary knowledge, including in a language unknown to the clinician and would provide the confidence that the prediction is being made for vocabulary skills for the language in question. Given the lack of measured vocabulary in these languages, a proxy for vocabulary knowledge that can be used is proficiency self-rating. Self-ratings have been shown to correlate with measures of proficiency in both Icelandic and English in a language-specific way (e.g., Elin Thordardottir, 2021) and self-rating differences between monolingual and bilingual groups have accurately predicted group differences in semantic fluency performance (Paap & Liu, 2014). In this thesis, we have added the novel measure of total productions in an unknown language on a semantic fluency test, which was assessed as a predictor self-rated proficiency.

Further, previous studies have focused on certain measures, such as total number of correct words, and have done so by following traditional scoring methods that do not allow for certain item variations (animal name variations for sex/gender) or superordinate category items. The purpose of verbal fluency tasks is to assess whether individuals can quickly produce words, and it is argued here that, since these variations are additional lexical items, they may be important in the search for an estimate of vocabulary size. In addition, in a clinical application, a simpler procedure is more useful as it is less time-consuming and less prone to inter-judge differences. Allowing for these variations is typically not assessed, and it is therefore unclear how this modified scoring might affect results. Since morphological variations are also not typically counted as correct answers, it is also unknown how counting such responses might have an effect. These modified measures were included in the current study.

Other measures, such as clustering and switching, have also been assessed in previous studies, although primarily with the purpose of assessing executive functioning abilities. These measures warrant exploration as predictors of vocabulary knowledge. Further, additional measures that may be more closely related to vocabulary knowledge and that focus on which specific words are produced are also worth exploring. This includes examining not only the number of words produced, but the lexical composition of the words produced in terms of word frequency, such as examining whether producing infrequent animal names tells us anything about vocabulary knowledge. Preliminary data showed that common words (e.g., cat, dog) were produced across many participants, whereas others were produced by very few participants (based on data from Dubé & Elin Thordardottir, 2019). This is consistent with research that shows language samples containing a small list of common words, with an even smaller list of low-frequency words (Richards & Malvern, 1997, as cited in Elin Thordardottir & Ellis Weismer, 2001). Therefore, beyond the number of words generated, this study specifically assessed which words were generated. It was of interest whether the production of infrequently produced words might indicate a larger vocabulary, which has indeed been shown in research involving language sample analysis (Rondal, 2003, as cited in Elin Thordardottir, 2016).

Another measure relating to lexical composition that has not been previously assessed is conceptual productions in participants who complete the verbal fluency task in two languages. Again, this includes focusing on which words are produced rather than strictly how many. In studies of bilingual vocabulary, unlike language-specific or total vocabulary which do not match monolingual norms, conceptual vocabulary size, which only counts overlapping words once, comes closer to doing so, particularly for children with unbalanced exposure to each language (e.g., Junker & Stockman, 2002; Elin Thordardottir, Rothenburg, Rivard, & Naves, 2006; Pearson et al., 1993). These overlapping words are also referred to as translation equivalents. Any words that are not overlapping between the two languages are uniquely produced in one of the languages. In this thesis, we included the measure of conceptual productions and translation equivalents produced on a verbal fluency task for bilingual participants to explore how overlap and uniqueness vary between languages in people with different amounts of exposure.

Finally, age effects have been documented, showing performance increasing with age (e.g., Brandeker & Elin Thordardottir, under review; Filippetti & Allegri, 2011; Friesen et al., 2015; Hurks et al., 2010) before a decline in late adulthood (Taler et al., 2019). However, longitudinal effects within participants also warrant investigation. In this thesis, we take advantage of the fact that some participants were tested twice to explore longitudinal effects.

Aims of Study

The specific aims of the study are the following.

1) To determine whether traditional and modified verbal fluency measures (semantic and phonemic) can be used to estimate vocabulary size in three languages for which we have vocabulary measures (English, French, Icelandic), using both same-language and cross-language estimations. It was hypothesized that verbal fluency measures would successfully predict vocabulary size in that same language but not in the other language. Relative predictive abilities of the different measures and performance in each language were explored.

2) To determine whether the number of productions on a semantic fluency task in languages unknown to the examiner can be used to estimate self-rated language proficiency. It was hypothesized that this prediction would be shown by total number of words produced predicting self-rated language performance.
3) To determine whether producing infrequent words on a semantic fluency task is indicative of higher vocabulary knowledge. It was hypothesized that producing infrequent words would generally indicate high receptive vocabulary scores; however, it was not known to what degree or whether producing infrequent words would provide a useful clinical measure.
4) To assess the use of translation equivalents in two languages on a semantic fluency task. This question was explored across different exposure groups. It was predicted that a greater proportion of translation equivalents would be produced on trials in the non-dominant language.
5) To determine whether performance on a semantic fluency task can indicate language dominance, operationalized as the language in which the most relative exposure has been received. It was predicted that language dominance could be determined in cases of clear dominance, but that in certain sub-samples (Icelandic trilinguals), performance would be more equally distributed across languages, as found in another study (Elin Thordardottir, 2021).
6) To describe verbal fluency performance over time (on average 4.69 years later; $n = 8$). Performance was expected to improve over time, given evidence of age effects (e.g., Brandeker & Elin Thordardottir, under review; Filippetti & Allegri, 2011; Friesen et al., 2015; Hurks et al., 2010).

Method

Design

The use of verbal fluency tasks to estimate vocabulary knowledge was assessed in a series of analyses. Testing took place in Montreal, Canada and Reykjavik, Iceland, allowing for exploration in two different linguistic contexts. Analyses 1 and 2 involved the entire sample from both testing locations. Analyses 3 to 6 involved different subgroups of the larger sample. Overall, these analyses assessed whether a variety of verbal fluency measures could serve as practical estimates of vocabulary size and indicators of language dominance across different ages and in different languages. All analyses but Analysis 3 investigated this in languages for which measured vocabulary scores were obtained, involving languages spoken by the experimenters. Traditional verbal fluency measures (total correct words, mean cluster size, number of switches) were assessed in addition to new measures of verbal fluency (modified scoring of total correct words, translation equivalents, use of infrequent words) in order to determine which measures might be most valuable for estimating vocabulary size. In Analysis 3, we innovatively attempted to use verbal fluency tasks for vocabulary estimation in languages unfamiliar to the experimenter, to examine whether this task can be used clinically even when the clinician does not know the language being assessed. Adequate estimation of vocabulary proficiency by any of these measures would support the proposal that verbal fluency can serve as a valuable clinical tool for this purpose when other tests are not available.

Ethics Approval

This project used data from studies that were approved by the Institutional Review Board of the Faculty of Medicine of McGill University and, for data collected in Iceland, also by the

Data Production Authority of Iceland (Persónuvernd). The parents of the participants signed an informed consent form and participants provided assent.

Participants

A total of 193 children and adolescents, ranging in age from 6 to 17 years, participated in this study. Participants in Montreal (n = 113) and in Iceland (n = 80) were recruited via invitation letters sent through their schools. Some participants in Montreal had previously participated in another study or an earlier version of this study and were called back to participate again.

Montreal Participants

A total of 113 individuals living in Montreal, ranging in age from 6 to 17 years old (M = 10.07, SD = 3.09) participated in this study. All participants had knowledge of French and attended French-language schools. Having French schooling in common, they all shared some exposure to French but varied based on languages spoken in the home and the community. Relative cumulative lifetime exposure to each language was determined from parent questionnaires, described below. Twenty-nine participants were considered monolingual speakers of French given their lack of sufficient exposure to a second language (exposed to French at least 90% of the time over their lifetime and inability to be tested in English). An additional 20 monolinguals were considered functionally monolingual, given their ability to complete tasks in both French and English. Forty-five participants were bilingual speakers of French and English, varying in their relative proficiency in each (more French: 61-89% lifetime exposure to French, balanced: 40-60% exposure to French, more English: <40% exposure to French). The remaining 18 participants were bilingual speakers of French and another language with no to little exposure to English. Participants with 10% or more exposure to a third language were excluded from analyses. A subset of participants (n = 10) were assessed twice, the second time being on average 4.69 years later (SD = 1.25). Unless otherwise specified, data from their second test sessions were used in analyses. In the longitudinal analysis, monolinguals who participated in the longitudinal portion of the study and who only completed the tasks in French at their first time of participation completed the tasks in both languages when re-tested later, given subsequent exposure to English in school.

Reykjavik Participants

Eighty participants living in Iceland were tested. Participants ranged in age from 10-16 years at the time of testing, M = 12.93, SD = 1.67 (exact age data could not be computed given inaccessible data in the context of the COVID-19 pandemic; n = 9). All participants attended school in Icelandic. Having Icelandic schooling in common, they all shared some exposure to Icelandic but varied in their exposure to languages outside of school. Thirty-three of these participants were L1 speakers of Icelandic, who learned English as a second language in school and through incidental exposure principally through the media. These participants were recruited as L1 speakers of Icelandic; however, in adolescence, most L1 speakers have acquired considerable English proficiency even though they do not use English in daily activities. The remaining participants spoke a language other than Icelandic or English at home. For the same reason that L1 speakers can be seen as bilinguals, these participants can be seen as trilinguals. Thirteen of them were participants who spoke Icelandic and another language at home, and whose third language was English with similar exposure as the bilingual L1 speakers. Thirty-two of these participants were L2 speakers of Icelandic, who spoke a variety of L1s at home and whose third language was English through incidental exposure. These Icelandic trilinguals varied in their relative exposure to Icelandic (Low: <40% lifetime exposure, Medium: 40-60%, High: >60% exposure). There are two participants for whom this information was not available at the time of writing and who were not included in analyses that required this information. In addition to Icelandic and English, the languages spoken by participants included the following: Albanian, Arabic, Bulgarian, Danish, German, Indonesian, Italian, Lithuanian, Norwegian, Polish, Portuguese, Russian, Serbian, Spanish, Thai, Vietnamese, Yoruba.

Procedure

Participants were tested as part of large-scale projects (Elin Thordardottir, 2019, 2021), part of which assessed verbal fluency and vocabulary knowledge (see also Brandeker & Elin Thordardottir, under review).

Montreal Participants

For speakers of English and French, testing was completed in English and in French during two separate testing sessions, with different experimenters. Monolingual French speakers only completed the tasks in French, unless they had sufficient knowledge of English to complete the tasks in both languages, as described above. Two types of verbal fluency tasks were administered: semantic fluency and phonemic fluency. The category used to assess semantic fluency was animals, which has been frequently used in the literature (Tombaugh et al., 1999) and is considered culturally neutral. The sounds used for the phonemic fluency task were /f/, /a/, /s/. These were chosen as they have been used in previous research (e.g., Rosselli et al., 2000). Only /f/ and /s/ were included in analyses because /a/ was judged to be pronounced differently by experimenters across languages and participants, thus these data were not considered to be collected in the same way consistently. Although explicit restrictions have been provided to participants in previous studies (no proper names, no morphological variations of the same root; e.g., Troyer, 2000; Troyer et al., 1997), no such restrictions were given in the current study given the young age of some participants (see Brandeker & Elin Thordardottir; under review). For each category, participants were given 60 seconds to produce as many words as possible, a time limit which has been imposed in previous studies (e.g., Rosselli et al., 2000; Troyer, 2000). From these tasks, multiple measures of verbal fluency were computed, described below.

Vocabulary knowledge was measured using formal vocabulary assessments in English and in French. To assess receptive vocabulary, participants were administered the Peabody Picture Vocabulary Test – Third Edition (PPVT-III; Dunn & Dunn, 1997) for English, and its French equivalent the *Échelle de vocabulaire en images Peabody* (EVIP, Dunn et al., 1993). These tests were administered following guidelines from the respective test manuals. Participants were shown images placed in 2x2 matrices and were asked to identify the object that best represented the word the experimenter provided. The test continued until participants reached the established stopping point (Dunn & Dunn, 1997; Dunn, et al., 1993). To assess expressive vocabulary, younger participants tested for a previous study (Brandeker & Elin Thordardottir, under review; *n* = 77) were assessed using the Expressive Vocabulary subtest from the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4; Semel, Wiig & Secord, 2003) in English, and from the *Évaluation clinique des notions langagières fondamentales* in French (CELF-CDN-F; Semel et al., 2009). Participants were asked to respond to the experimenter's questions (e.g., What is this?). For older, newly recruited participants (*n* = 36), English expressive vocabulary was assessed with the Expressive Vocabulary Test – Second Edition (EVT; Williams, 1997). Participants were shown pictures one at a time and asked to use one word to answer the experimenter's question about the picture (e.g., What is this?, what is she doing?). No formal French expressive vocabulary test was available for older participants.

Reykjavik Participants

Testing was done as part of a larger protocol of tests in Icelandic. Verbal fluency was the only test in the session that also focused on English and the participants' home language in cases in which this was not Icelandic. For the verbal fluency tasks, a similar procedure was followed as with the Montreal participants, in English and in Icelandic. For their verbal fluency tasks in Icelandic, participants were assessed on both semantic and phonemic fluency. All participants also completed semantic (animal) fluency in English. Participants who spoke a language other than Icelandic at home (n = 36) were additionally assessed for semantic fluency in that language. The only difference between the methods used in Montreal and in Iceland was that for phonemic fluency, participants were asked to provide words that began with the given letters (F, S), rather than the sounds. However, words beginning with these sounds (/f, s/) always begin with these letters in Icelandic, unlike in English (e.g. phenomenon vs. fantastic), so this was not believed to be of importance. Again, unlike in previous studies (e.g., Troyer, 1997), no specific restrictions (no proper names, no morphological variations of the same root) beyond the main criteria were provided.

Icelandic receptive and expressive vocabulary were assessed by a test designed in Icelandic, Milli mála (Elin Thordardottir, 2011), which provides separate receptive and expressive scores. In this study, testing of English vocabulary could not be done due to time restrictions and testing of vocabulary in the home languages could not be done due to lack of tests and testers who spoke these languages. Instead, participants were asked to self-evaluate their understanding and speaking proficiencies in each of their languages on a 4-point scale: 1= very well, 2= well, 3= fairly well, 4= not at all (*mjög vel, vel, sæmilega, alls ekki*). This self-rating scale has been used in a previous study with another sample of adolescents tested in

Icelandic and English with correlations between tested and self-rated scores supporting the language-specific validity of the self-ratings (Elin Thordardottir, 2021).

For all participants, questionnaires were completed by participants' parents for the purpose of gathering data about their language exposure over their lifetime. Answers from these questionnaires were used to determine relative exposure to their languages over their lifetime (Montreal: English, French; Iceland: Icelandic), similar to the methods used in studies in Quebec and in Iceland by Elin Thordardottir (2011; 2019; 2021). Relative exposure was then used as a proxy for proficiency in order to designate languages as the dominant vs. non-dominant language given previous evidence of a strong relationship between cumulative exposure and proficiency (e.g., Elin Thordardottir, 2011; 2019; Hoff et al., 2012).

Scoring Verbal Fluency Tasks

Data were transcribed from audio or video recordings. A subset of participants (younger children in Montreal, n = 77) were tested for an earlier study (Brandeker & Elin Thordardottir, under review). Their data had therefore already been transcribed and scored, and these previously scored measures were used for the current thesis. This was the case for measures of mean cluster size and number of switches. However, given the inclusion of modified measures of total correct words, the traditional measure of total correct words was re-scored for the current thesis, in order to ensure consistency in scoring procedures. All other novel measures and all data for the subsequently recruited participants in Montreal were scored by the author for the purpose of this thesis.

For the Icelandic participants, English trials were transcribed and scored by the author. Icelandic trials were transcribed and translated to English by an Icelandic-English speaking research assistant, and scoring was then completed by the author, with assistance from speakers of Icelandic as needed. Word productions in a home language other than Icelandic were counted by an Icelandic research assistant to ensure Icelandic productions during the task were not given credit. Figure 1 summarizes the measures scored from the verbal fluency tasks, described in further detail below.

Figure 1

Verbal Fluency Measures Included in Thesis

Measure	Replication	Novel Measure
Semantic Fluency		
Total words: Traditional scoring	✓	
Total words: Modified scoring		✓
Mean cluster size	✓	
Number of switches	✓	
Translation Equivalents*		✓
Production of infrequent words*		✓
Total productions in language unknown to tester**		✓
Phonemic Fluency		
Total words: Traditional scoring	✓	
Total words: Modified scoring		✓
Mean cluster size	✓	
Number of switches	✓	

*For bilingual speakers
**For speakers of a third language

Total Correct Words

Traditional Measure of Total Correct Words. Verbal fluency tasks were scored using the same procedure in each language. For the total number of correct words measure, following the traditional method of scoring (e.g., Troyer et al., 1997), only words that adhered to the given category in the appropriate language were counted. Therefore, spontaneous codeswitches (words produced in another language) were coded as such and were not included in the total. Repetitions of words were coded as repetitions without being included in this measure. In previous studies, versions of animals that differed in sex (e.g. hen/rooster) and age (e.g. cat/kitten) were counted as a single correctly produced word (e.g., Troyer et al., 1997). Further, superordinate categories (e.g. bird) were not counted when more specific category items were provided (e.g. robin, eagle). Further, for phonemic fluency in previous studies, different words with the same root (e.g. run/running) were counted as one single correct word. Responses were first scored using these guidelines.

Modified Measure of Total Correct Words. Modified measures of total number of correct words were also computed. For semantic fluency, this modified measure allowed for superordinate categories and sex/age variants of animal names as correct words. For example, if both "bird" and "parrot" were produced, under this modified measure, both were scored as correct words. Further, if "hen" and "rooster" or "dog" and "puppy", respectively, were both produced, both were scored as correct words. A modified phonemic fluency measure was also computed, allowing for productions of the same root word with different morphological derivations. For example, if "ran" and "running" were produced, both were scored as correct words. The traditional and modified measures were used in separate analyses.

Mean Cluster Size

As mentioned, a subset of the scores for the measure of mean cluster size (*n* = 77) were taken from a previous study (Brandeker & Elin Thordardottir, under review). The remaining participants were scored using the same specific criteria, which followed a procedure based on previous research (e.g., Troyer et al., 1997). Within the semantic and the phonemic tasks, clusters can be formed based on both semantic and phonemic attributes. First, for the semantic category “animals”, items can be clustered based on four semantic attributes: habitat, zoological family, family members, and human use. Second, for the phonemic fluency tasks, semantic clusters can be formed using the following criteria: super-ordinates or subordinates, words within the same semantic category (snake & serpent), words with a close semantic or contextual relation (friends & family). Finally, categories from both the semantic and phonemic fluency tasks can be clustered under the following phonemic attributes: words that rhyme, words that begin with the same two phonemes, words that differ by only one vowel, and homonyms (when made clear by the participant that these are not repetitions of the same item). Following these criteria, it is possible for multiple clusters to exist within one larger cluster. When this occurred, only the larger cluster was considered in the measure of cluster size, as done previously (e.g., Brandeker & Elin Thordardottir, under review; Troyer, 2000). Clusters can also overlap, when a word belongs to two consecutive clusters, in which case such a word was included in the measure for both clusters. Note that cluster size is counted starting with the second word in the cluster such that a single-word cluster has a cluster size of 0, two-word clusters have a cluster size of 1, three-word clusters have a size of 2, and so on. All productions within the same semantic category (animals) or same phonemic attribute (F, S) were averaged to calculate the measure of mean cluster size, in a given language.

Number of Switches

Subsequently, the mean number of switches (between clusters) was calculated by counting the number of times participants switched from one cluster to the next (i.e., number of clusters – 1).

Special Considerations

Variables for semantic fluency were computed by using the scores obtained for the category “animals”. Variables for phonemic fluency were computed by adding the scores from the different trials (/f, s/) and dividing by 2, for an average score. This is consistent with previous research (Bialystok et al., 2008; Brandeker & Elin Thordardottir, under review; Friesen et al., 2015). It should be noted that, typically in studies of phonemic fluency, participants are asked to produce words that begin with a certain letter. Since some participants (younger participants tested in Montreal) were pre-literate at the time of testing, they were asked to produce words that began with a certain sound. In order to maintain the same procedure across participants, the same instructions were provided for all participants completing phonemic fluency tasks in Montreal. However, given the older participants’ reading abilities, it is possible that they relied more heavily on orthographic knowledge. In order to account for this, when measuring the total number of correct words produced, words that either began with the correct sound or letter (or oftentimes both) were counted as correct and were coded accordingly in order to keep track. This only affects English and French, such that words that start with the sound /f/ can start with the letters F or PH, and words that start with the sound /s/ can start with the letters S or C. Words that started with the correct letter but not sound (e.g., shoe) were rarely produced. As mentioned, in Icelandic, words starting with /f/ and /s/ are unambiguous in terms of letter or sound. There is no other way to spell these sounds and these letters cannot be pronounced any other way in word-initial position.

Translation Equivalents

For Analysis 4, the number of cross-language synonyms, or translation equivalents, produced was counted for bilingual participants. This type of measure has not previously been included in verbal fluency studies. This was done for example for a participant tested in English and French, who produced both 'horse' and '*cheval',* its translation, by counting these only as one production. The proportion of translation equivalents/total productions in a given language was then computed to be used in analyses, providing information on the proportion of words in one language (e.g., French) for which the translation equivalent was also produced in the other language (e.g., English), and subsequently on the proportion of words uniquely produced in each language. This information could indicate the relative contributions of vocabulary knowledge in each language to conceptual vocabulary size.

Scoring in an Unknown Language

A subset of participants in Iceland (n = 36) completed the semantic fluency task in an additional language. Only a measure of total productions was scored for these. Each production was awarded one point, regardless of whether the word conformed to the category or was repeated, since the experimenter did not know the language. This score was computed by a native speaker of Icelandic, who could differentiate between word productions in the unknown language and utterances in Icelandic directed to the experimenter.

Reliability Scoring

Reliability was calculated for 20% of participants at each testing location. Reliability coding was completed by a French-English speaker for the Canadian data and by Icelandic-English speakers for the Icelandic data, all blind to the hypotheses of the study. Reliability scores were then computed by the author by comparing the author's original scores to the reliability scorers' scores, using the following formula: ((Total scores Author + Reliability Scorer) – Amount of Discrepancy between total scores of Author & Reliability Scorer)/Total scores Author + Reliability Scorer. For data scored for an earlier study from our lab (a subset of scores for mean cluster size and number of switches), see Brandeker and Elin Thordardottir (under review) for reliability scores. Recall, for these participants (*n* = 77), that total correct words and novel measures for the entire sample were re-scored/scored by the first author for the current thesis. Reliability for 20% of these data was high across measures for English semantic fluency (total words: 98.90% total words modified 99.27%) and phonemic fluency (total words: 100.00%, total words modified 100.00%), and for French semantic fluency (total words: 98.65% total words modified 99.18%) and phonemic fluency (total words: 98.58% total words modified 99.38%). Reliability for translation equivalents on the semantic fluency task was also high (98.56%). For reliability scores for 20% of newly recruited participants (*n* = 116), see Table 1.

Table 1

Reliability Scores

	Semantic Fluency			
	Canada		Iceland	
Measures	**English**	**French**	**English**	**Icelandic**
Transcription	98.63%	99.22%	99.52%	95.99%
Total Words	99.85%	100.00%	98.67%	97.80%
Total Words: Modified	99.72%	99.72%	99.00%	97.27%
Total Cluster Size	96.58%	96.12%	93.75%	91.08%
Number of Switches	93.67%	94.61%	91.19%	91.45%
Number of translation equivalents	96.73%		95.49%	

	Phonemic Fluency			
Transcription	93.75%	93.59%	N/A	95.26%
Total Words	97.57%	98.77%	N/A	97.13%
Total Words: Modified	98.51%	99.02%	N/A	97.57%
Total Cluster Size	89.17%	91.89%	N/A	82.40%
Number of Switches	93.75%	93.33%	N/A	92.37%

Note that mean cluster size is influenced by the number of clusters and consequently by the number of switches. In order to avoid having reliability outcomes from the number of switches affect our calculation of cluster size reliability, total cluster size prior to calculating the average was reported. Reliability was high across measures, testing locations, and languages.

Results

Statistical Approach

Given the exploratory nature of this unique study, numerous statistical analyses were needed to determine whether each of several verbal fluency task scores could predict scores on different vocabulary tests. In addition, both same-language and cross-language regressions were needed to assess the potential language-specificity of regressions. The risk of false positives was a reality. However, a stringent Bonferonni correction would have inflated the risk of false negatives given a loss of power (e.g., Armstrong, 2014), in a context in which finding these negatives is of theoretical interest. Further, group comparisons, specifically those including language dominance, included multiple *t*-tests done separately for different language and exposure groups. ANOVAs were deemed unfit for these questions since opposite effects of dominance would cancel each other out in the omnibus test, whereas the tests of real interest would have remained the paired-samples *t*-tests, whether performed as post hocs or as planned analyses. Partly offsetting the problem of multiple statistical tests in the presence of replication within the study, whereby the ability of verbal fluency to predict vocabulary size was tested in two independent samples of participants, in Montreal and in Reykjavik. Further replication will be required to confirm the findings of the study.

Analysis 1: Semantic Fluency

English Semantic Fluency

The main purpose of this analysis was to determine whether semantic fluency measures can predict measured vocabulary size in a sample of individuals across a wide age range and with different linguistic backgrounds. Montreal participants' performance on the semantic fluency task of naming animals in English was assessed as a predictor of receptive and expressive vocabulary knowledge in English and of receptive vocabulary in French using simple linear regressions. Assumptions for linear regressions were first tested and mostly met. Only the assumption of homoscedasticity was violated in a few cases. Note that it has been argued that when the main purpose is to predict the value of the dependent variable, as was the case here, violation of homoscedasticity may not be an issue (Frost, 2019). Further, findings with heteroscedastic data were replicated using scores from other vocabulary tests for which the data showed homoscedasticity (i.e., the heteroscedastic regressions on the English task were replicated by homoscedastic regressions using other English vocabulary scores). Therefore, analyses were run as planned (See Tables 1 and 2).

Total correct words in English was assessed as a predictor variable. The result was significant for all English vocabulary tests: PPVT, CELF, EVT. The result was not significant for either French test: EVIP, CELF-CDN-F.

When the modified measure of total correct words in English was assessed as the predictor variable, the result was significant for all English tests: PPVT, CELF, and EVT. The result was not significant either French test: EVIP, CELF-CDN-F.

When mean cluster size in English was assessed as the predictor, the result was significant for PPVT scores and EVT scores. The result was not significant for either French test (EVIP, CELF-CDN-F) or for the CELF.

When number of switches on the English task was assessed as the predictor, results were significant for PPVT scores, EVIP scores, CELF scores, and EVT scores. Results were not significant for CELF-CDN-F scores. See Tables 1 and 2 for summaries of results, separated by same vs. cross-language regressions. Rows highlighted in yellow show *p* values <.05 and rows highlighted in green show *p* values <.01.

Table 1

Same-Language Regressions with English Semantic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	PPVT	.43	.42	46.52 (1,62)	<.001	.66
	CELF	.37	.34	13.85 (1, 24)	.001	61
	EVT	.60	.60	53.67 (1, 36)	< .001	.77
Total correct words *(modified scoring)*	PPVT	.49	.48	58.68 (1,62)	< .001	.70
	CELF	.43	.40	17.90 (1, 24)	< .001	.65
	EVT	.70	.70	83.14 (1, 36)	< .001	.84
Mean cluster size	PPVT	.07	.06	4.78 (1, 62)	.032	.27
	CELF	.02	-.02	.61 (1, 24)	.442	.16
	EVT	.13	.10	5.24 (1, 36)	.028	.36
Number of switches	PPVT	.26	.25	21.66 (1, 62)	< .001	.51
	CELF	.24	.21	7.59 (1, 24)	.011	.49
	EVT	.39	.38	23.24 (1, 36)	< .001	.63

Note: Receptive vocabulary test: PPVT; Expressive vocabulary tests: CELF (expressive vocabulary subtest); EVT

Table 2

Cross-language Regressions with English Semantic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words	EVIP	.06	.04	3.25 (1, 56)	.077	.23
(traditional scoring)	CELF-CDN-F	.13	.10	3.67 (1, 24)	.068	-.36
Total correct words	EVIP	.06	.04	3.53 (1, 56)	.065	.24
(modified scoring)	CELF-CDN-F	.11	.08	3.06 (1, 24)	.093	-.34
Mean cluster size	EVIP	<.01	-.01	.25 (1, 56)	.618	-.07
	CELF-CDN-F	.06	.02	1.61 (1, 24)	.216	-.25
Number of switches	EVIP	.09	.07	5.44 (1,56)	.023	.30
	CELF-CDN-F	.01	-.04	.13 (1, 24)	.719	.07

Note: Receptive test: EVIP; Expressive test: CELF-CDN-F (expressive vocabulary subtest)

French Semantic Fluency

When total correct words produced in French was assessed as the predictor, results were significant for PPVT scores and scores on both French tests (EVIP, CELF-CDN-F). Results were not significant for either English expressive test: CELF, EVT.

When the modified measure of total correct words in French was assessed as the predictor, results were significant for PPVT scores and scores on both French vocabulary tests, (the EVIP and the CELF-CDN-F). Results were not significant for scores on either English expressive test: CELF, EVT.

When mean cluster size in French was assessed as the predictor variable, results were significant for EVIP scores. Results were not significant for scores on the PPVT, CELF, CELF-CDN-F, or EVT.

When number of switches on the French task was assessed as the predictor, results were significant for scores on receptive vocabulary tests: PPVT, EVIP. Results were not significant for expressive vocabulary scores: CELF, CELF-CDN-F, EVT.

Table 3

Same -Language Regressions with French Semantic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	EVIP	.36	.35	55.68 (1, 99)	< .001	.60
	CELF-CDN-F	.13	.12	9.14 (1, 61)	.004	.36
Total correct words *(modified scoring)*	EVIP	.39	.38	62.80 (1, 99)	< .001	.62
	CELF-CDN-F	.14	.12	9.72 (1, 61)	.003	.37
Mean cluster size	EVIP	.05	.04	5.37 (1, 100)	.023	.23
	CELF-CDN-F	.05	.03	3.20 (1, 62)	.079	.22
Number of switches	EVIP	.25	.24	32.74 (1, 100)	< .001	.50
	CELF-CDN-F	.03	.02	1.97 (1, 62)	.165	.18

Note: Receptive test: EVIP; Expressive test: CELF-CDN-F (expressive vocabulary subtest)

Table 4

Cross-Language Regressions with French Semantic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	PPVT	.12	.10	7.53 (1, 58)	.008	.34
	CELF	.14	.10	3.52 (1, 22)	.074	-.37
	EVT	.02	-.01	.55 (1, 33)	.465	-.13
Total correct words *(modified scoring)*	PPVT	.15	.14	10.27 (1,58)	.002	.39
	CELF	.09	.04	2.06 (1, 22)	.166	-.29
	EVT	.01	-.02	.36 (1, 33)	.550	-.10
Mean cluster size	PPVT	<.01	-.01	.09 (1, 59)	.760	.04
	CELF	.05	.01	1.22 (1, 23)	.280	-.26
	EVT	.02	-.01	.70 (1, 33)	.409	-.14
Number of switches	PPVT	.14	.12	9.29 (1, 59)	.003	.37
	CELF	.03	-.01	.69 (1, 23)	.417	-.17
	EVT	<.01	-.03	.05 (1, 33)	.829	.04

Note: Receptive vocabulary test: PPVT; Expressive vocabulary tests: CELF (expressive vocabulary subtest); EVT

Icelandic Semantic Fluency

Replication of the Montreal results was attempted with results from the Icelandic participants who completed both verbal fluency and vocabulary testing (*n* = 78) by assessing performance on semantic fluency in Icelandic as a predictor of Icelandic receptive and expressive vocabulary. Assumptions were first tested and mostly met. Again, some regressions did not meet the assumption of homoscedasticity. However, this was not considered to be problematic for this particular research question.

When total correct words produced in Icelandic was assessed as the predictor variable, results were significant for receptive vocabulary scores and expressive vocabulary scores.

When the modified measure of total correct words produced in Icelandic was assessed as the predictor, results were significant for receptive vocabulary scores and expressive vocabulary scores.

When mean cluster size in Icelandic was assessed as the predictor variable, results were not significant for receptive vocabulary scores or expressive vocabulary scores.

When number of switches in Icelandic was assessed as the predictor, results were significant for receptive vocabulary scores and expressive vocabulary scores.

Table 5

Same-Language Regressions with Icelandic Semantic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words	Receptive	.51	.50	78.64 (1, 76)	< .001	.71
(traditional scoring)	Expressive	.42	.42	55.95 (1, 76)	< .001	.65
Total correct words	Receptive	.49	.48	71.92 (1, 76)	< .001	.70
(modified scoring)	Expressive	.43	.42	56.65 (1, 76)	< .001	.65
Mean cluster size	Receptive	.02	.01	1.74 (1, 76)	.191	.15
	Expressive	<.01	-.01	.22 (1, 76)	.642	.05
Number of switches	Receptive	.12	.11	10.24 (1, 76)	.002	.35
	Expressive	.18	.16	16.13 (1, 76)	< .001	.42

Analysis 2: Phonemic Fluency

English Phonemic Fluency

These same predictions were then assessed using phonemic fluency measures to predict vocabulary scores. Note, phonemic fluency scores consisted of the averaged phonemic fluency scores from both trials ((F+S)/2). For participants missing one of these trials, only the available trial was used as their phonemic score. This question was first tested with the participants in Montreal, Canada. Assumptions were tested and mostly met. Again, in some cases, the assumption of homoscedasticity was violated, which was not believed to be problematic for these analyses.

When total correct words produced in English was assessed as the predictor variable, results were significant for scores on the PPVT, EVIP, CELF, and EVT. Results were not significant for scores on the CELF-CDN-F.

When the modified measure of total correct words in English was assessed as the predictor, results were significant for scores on the PPVT, EVIP, CELF, and EVT. The result was not significant for scores on the CELF-CDN-F.

When mean cluster size in English was assessed as the predictor, results were not significant for any vocabulary assessment.

When number of switches was assessed as the predictor, results were significant for scores on the PPVT, EVIP, and EVT. Results were not significant for scores on the CELF or CELF-CDN-F. See Tables 6 and 7 for summaries of regressions.

Table 6

Same-Language Regressions with English Phonemic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	PPVT	.49	.48	60.36 (1, 64)	< .001	.70
	CELF	.18	.14	5.07 (1, 24)	.034	.42
	EVT	.39	.37	24.11 (1, 38)	< .001	.62
Total correct words *(modified scoring)*	PPVT	.49	.48	61.62 (1, 64)	< .001	.70
	CELF	.16	.13	4.60 (1, 24)	.042	.40
	EVT	.40	.38	25.22 (1, 38)	< .001	.63
Mean cluster size	PPVT	.06	.04	3.80 (1, 64)	.056	.24
	CELF	<.01	-.04	.03 (1,24)	.847	.04
	EVT	.33	.31	18.51 (1, 38)	< .001	.57
Number of switches	PPVT	.41	.40	43.98 (1, 64)	< .001	.64
	CELF	.12	.09	3.41 (1, 24)	.077	.35
	EVT	.33	.31	18.51 (1, 38)	< .001	.57

Note: Receptive vocabulary test: PPVT; Expressive vocabulary tests: CELF (expressive vocabulary subtest); EVT

Table 7

Cross-Language Regressions with English Phonemic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	EVIP	.29	.28	23.33 (1, 58)	< .001	.54
	CELF-CDN-F	.06	.02	1.61 (1, 24)	.217	.25
Total correct words *(modified scoring)*	EVIP	.30	.29	24.95 (1, 58)	< .001	.55
	CELF-CDN-F	.08	.04	2.10 (1, 24)	.161	.28
Mean cluster size	EVIP	.03	.02	1.90 (1, 58)	.173	.18
	CELF-CDN	.01	-.03	.26 (1, 24)	.612	.10
Number of switches	EVIP	.27	.25	20.95 (1, 58)	< .001	.52
	CELF-CDN-F	.05	.01	1.30 (1, 24)	.265	.23

Note: Receptive test: EVIP; Expressive test: CELF-CDN-F (expressive vocabulary subtest)

French Phonemic Fluency

These same regressions were then completed using scores on the French verbal fluency tasks. When total correct words produced in French was assessed as the predictor, results were significant for scores on the PPVT, EVIP, and CELF-CDN-F. Results were not significant for scores on the CELF or EVT.

When the modified measure of total correct words in French was assessed as the predictor variable, results were significant for scores on the PPVT, EVIP, and CELF-CDN-F. Results were not significant for scores on the English expressive tests (CELF, EVT).

When mean cluster size in French was assessed as the predictor, the result was significant for EVIP scores. The results were not significant for any other scores: EVIP, CELF, CELF-CDN-F, EVT.

When number of switches in French was assessed as the predictor variable, results were significant for scores on receptive vocabulary tests: PPVT, EVIP. Results were not significant for scores on any expressive vocabulary tests: CELF, CELF-CDN-F, EVT.

Table 8

Same-Language Regressions with French Phonemic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	EVIP	.39	.39	64.83 (1, 101)	< .001	.63
	CELF-CDN-F	.07	.05	4.42 (1, 63)	.040	.26
Total correct words *(modified scoring)*	EVIP	.41	.41	70.72 (1, 101)	< .001	.64
	CELF-CDN-F	.08	.06	5.34 (1, 63)	.024	.28
Mean cluster size	EVIP	.04	.04	4.72 (1, 102)	.032	.21
	CELF-CDN	.04	.03	2.95 (1, 64)	.091	.21
Number of switches	EVIP	.35	.34	53.92 (1, 102)	< .001	.59
	CELF-CDN-F	.04	.02	2.32 (1, 64)	.133	.19

Note: Receptive test: EVIP; Expressive test: CELF-CDN-F (expressive vocabulary subtest)

Table 9

Cross-Language Regressions with French Phonemic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words *(traditional scoring)*	PPVT	.27	.26	21.86 (1, 60)	< .001	.52
	CELF	.01	-.03	.33 (1, 24)	.572	-.12
	EVT	<.01	-.03	.04 (1, 33)	.836	.04
Total correct words *(modified scoring)*	PPVT	.29	.28	24.16 (1, 60)	< .001	.54
	CELF	.01	-.03	.28 (1, 24)	.603	-.11
	EVT	<.01	-.03	.11 (1, 33)	.743	.06
Mean cluster size	PPVT	.01	-.01	.28 (1, 61)	.598	.07
	CELF	.10	.06	2.79 (1, 25)	.108	-.32
	EVT	<.01	-.03	.01 (1, 33)	.918	-.02
Number of switches	PPVT	.28	.27	24.11 (1, 61)	< .001	.53
	CELF	.01	-.03	.18 (1, 25)	.677	-.08
	EVT	<.01	-.03	.14 (1, 33)	.715	.06

Note: Receptive vocabulary test: PPVT; Expressive vocabulary tests: CELF (expressive vocabulary subtest); EVT

Icelandic Phonemic Fluency

Simple linear regressions were conducted with Icelandic verbal fluency measures as predictors of vocabulary scores. Assumptions were tested. All were met, except the assumption of homoscedasticity when number of switches was used to predict expressive vocabulary.

When total correct words produced in Icelandic was assessed as the predictor, results were significant for scores of receptive vocabulary and expressive vocabulary.

When the modified measure of total correct words produced in Icelandic was assessed as the predictor, results were significant for scores of receptive vocabulary and expressive vocabulary.

When mean cluster size in Icelandic was assessed as the predictor variable, results were not significant for either vocabulary measure.

When number of switches in Icelandic was assessed as the predictor, results were significant for scores of receptive vocabulary and expressive vocabulary.

Table 10

Same-Language Regressions with Icelandic Phonemic Fluency Measures

VF score predictor	*Test outcome measure*	R^2	*Adjusted* R^2	*F (df)*	*p*	β
Total correct words	Receptive	.48	.48	70.27 (1, 75)	<.001	.70
(traditional scoring)	Expressive	.34	.33	37.95 (1, 75)	<.001	.58
Total correct words	Receptive	.44	.43	59.18 (1, 75)	<.001	.66
(modified scoring)	Expressive	.28	.27	28.62 (1,75)	<.001	.53
Mean cluster size	Receptive	<.01	-.01	.37 (1, 75)	.545	.07
	Expressive	<.01	-.01	.18 (1, 75)	.676	.05
Number of switches	Receptive	.48	.47	68.00 (1, 75)	<.001	.69
	Expressive	.32	.31	35.69 (1, 75)	<.001	.57

Analysis 3: Estimating Proficiency in an Unknown Language

A subset of Icelandic participants also completed the animal fluency task in a home language unknown to the examiners (n = 36). Since we did not have measured vocabulary scores for these participants, verbal fluency (total productions) was assessed as a predictor for participants' self-ratings of their comprehension and production abilities in a given language on a 4-point scale (1: very well, 2: well, 3: fairly, 4: not at all), using simple linear regression. These ratings were assumed to be continuous and have successfully been used in correlation analyses in a previous study (Elin Thordardottir, 2021). Assumptions were tested and met. Results were significant for self-ratings of comprehension abilities (R^2= .304, $F(1, 34)$= 14.86, $p < .001$, *adjusted* R^2= . 284, β = -.55), and production abilities (R^2= .522, $F(1, 34)$= 37.18, $p < .001$, *adjusted* R^2= .508, β= -.72). In both cases, these regressions showed that the more items produced, the higher the probability of having a lower score on the scale (i.e. higher proficiency).

Recall that when this task was performed in Icelandic, comparable magnitudes of standard beta coefficients were obtained (receptive: .70; expressive: .65).

Because the modified measure of total correct words on the semantic fluency task yielded the best overall predictions of vocabulary size in the analyses reported thus far, and was completed by participants in all of their languages, all further analyses reported in the sections that follow used that measure.

Analysis 4: Naming Frequency and Translation Equivalents

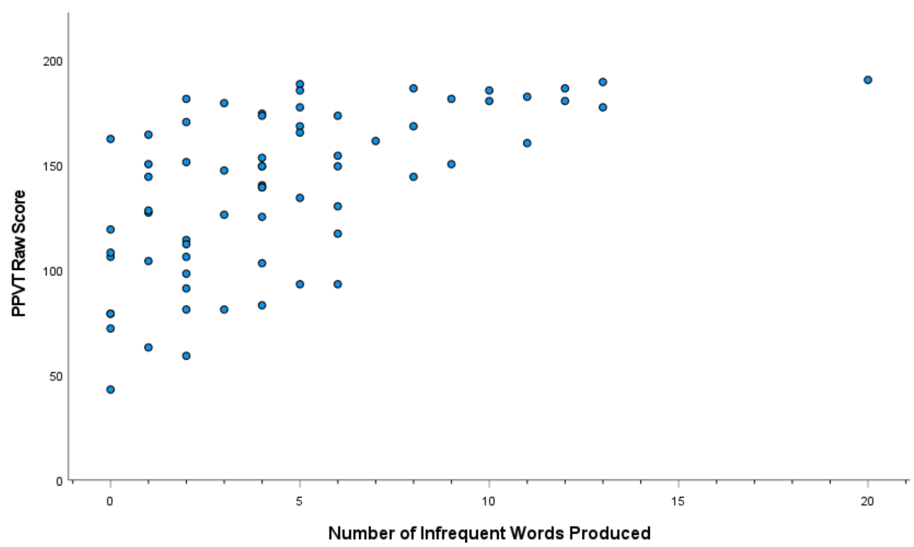
Naming Frequency

To address the frequency at which individual animal names were produced, rankings of the most-to-least frequently produced words were established at the group level in each language within each linguistic context (Montreal: French, English; Iceland: Icelandic). Words tied for the 10 least produced words were considered "infrequent". The number of infrequent words produced by each participant was then computed. Data from both time points for participants who participated twice (*n* = 10) were included in the determination of infrequent words and in the following graphs. See Appendices 1 to 3 for a list of infrequent words in each language.

English Naming Frequency. The use of infrequent words was studied as a potential indicator of high receptive vocabulary scores. Figures 2 and 3 are scatterplots showing vocabulary scores as a function of the number and the percentage, respectively, of infrequent words produced by individual participants. Based on these scatterplots, it appears that participants who said a greater number of infrequent words in English obtained higher scores on the PPVT. Clinically, it would be of use to establish a certain number that differentiates high and low vocabulary performers. Visual inspection of the graphs indicates that, as a rule of thumb, producing 10 infrequent words seemed to indicate likely having high PPVT scores. However, note that some participants with higher PPVT scores produced few infrequent words, therefore, high production of infrequent words may be a more valuable indicator clinically than low production of infrequent words. Given the exploratory nature of this study, percentage of infrequent words, in addition to raw number, was investigated as a potential indicator of high vocabulary knowledge. This measure did not provide such information, likely since this measures covaries with the number of words produced.

Figure 2

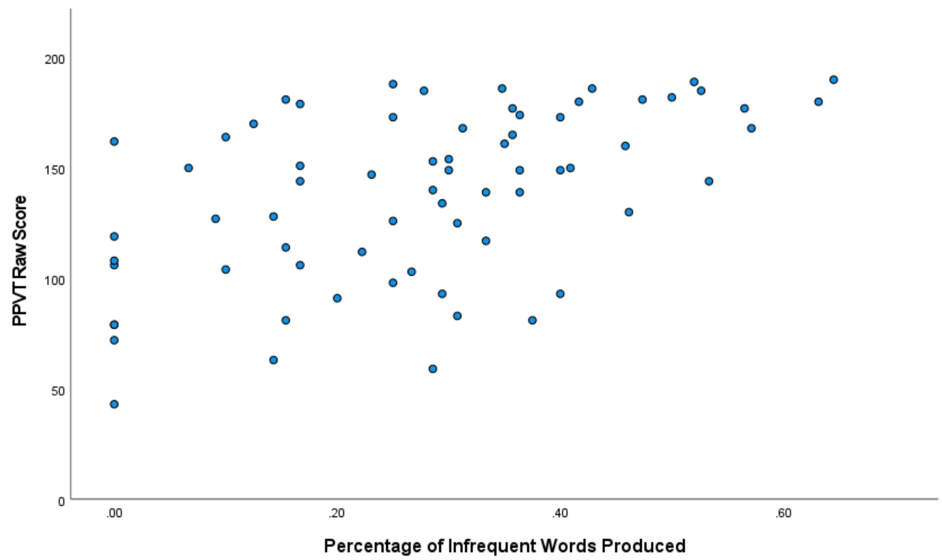
Scatterplot of PPVT Raw Scores by Number of Infrequent Words Produced



Note: *n* = 68

Figure 3

Scatterplot of PPVT Raw Scores by Percentage of Infrequent Words Produced

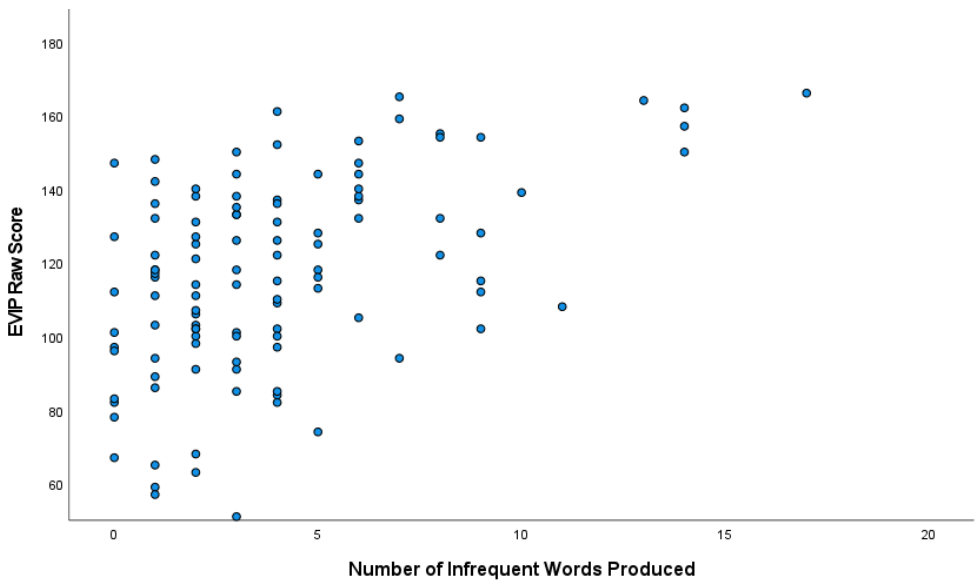


Note: n = 68

French Naming Frequency. Similarly to in English, producing more infrequent words in French appeared to indicate higher EVIP scores. Again, producing 10 words appeared to indicate high EVIP scores. Percentage of infrequent words produced in French was not indicative of EVIP scores. See Figures 4 and 5 for individual data points.

Figure 4

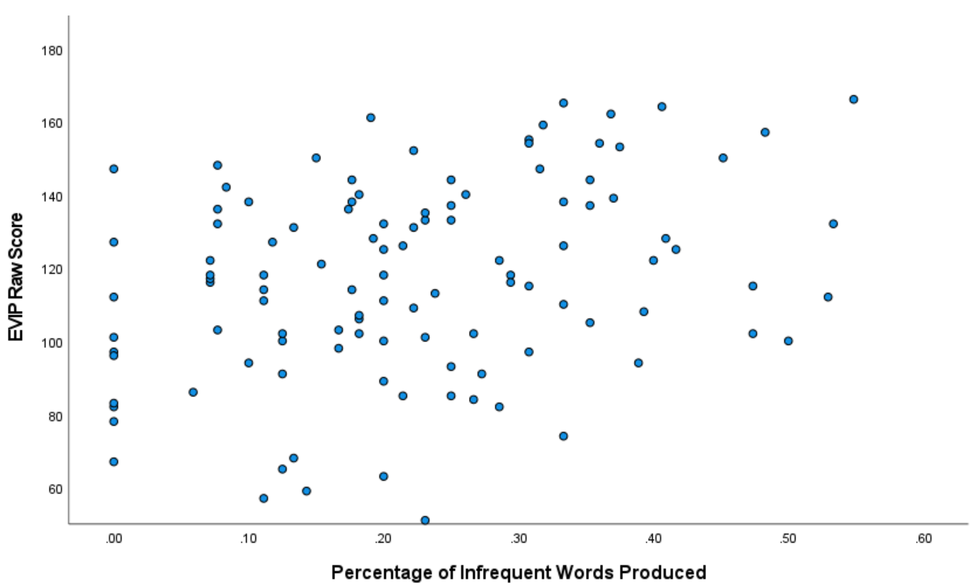
Scatterplot of EVIP Raw Scores by Number of Infrequent Words Produced



Note: n = 112

Figure 5

Scatterplot of EVIP Raw Scores by Percentage of Infrequent Words Produced

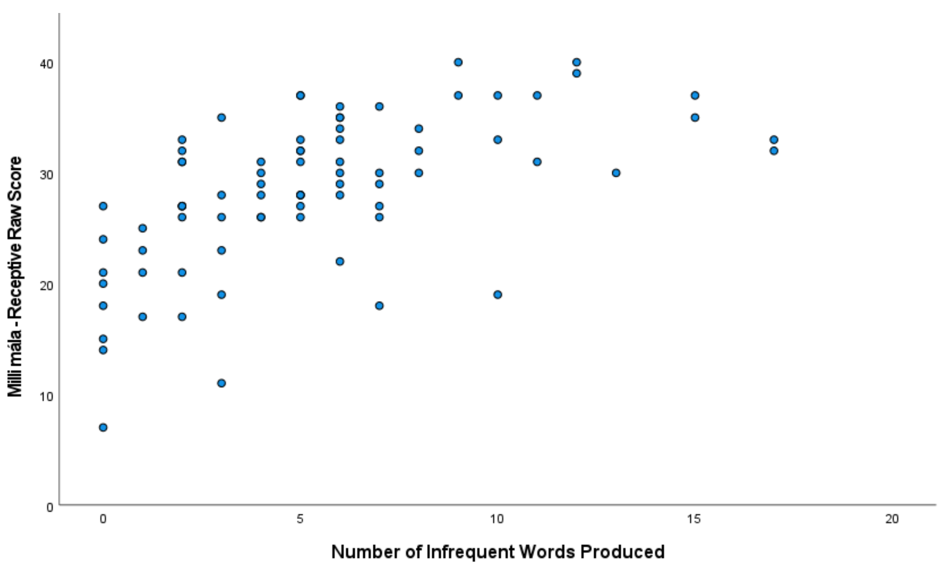


Note: *n* = 112

Icelandic Naming Frequency. Based on Icelandic data, it once again appeared that producing a higher number of infrequent words indicated higher receptive vocabulary scores, but percentage of infrequent words produced made no such indication. Once again, using 10 infrequent words as our rule of thumb, this appeared to indicate high receptive vocabulary scores. See Figures 6 and 7 for individual data points.

Figure 6

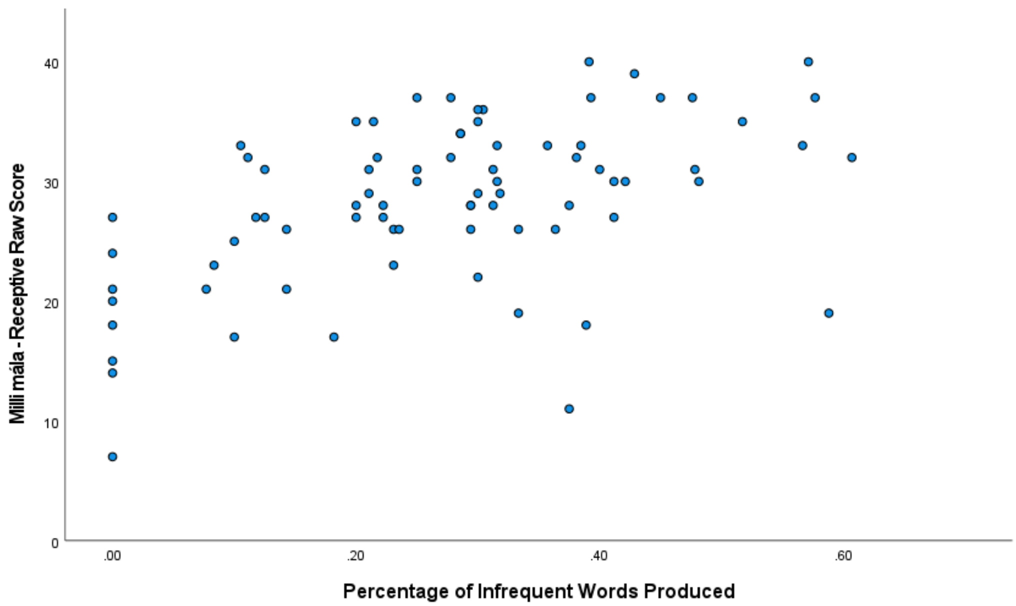
Scatterplot of Milli Mála Raw Scores by Number of Infrequent Words Produced



Note: *n* = 78

Figure 7

Scatterplot of Milli Mála Raw Scores by Percentage of Infrequent Words Produced



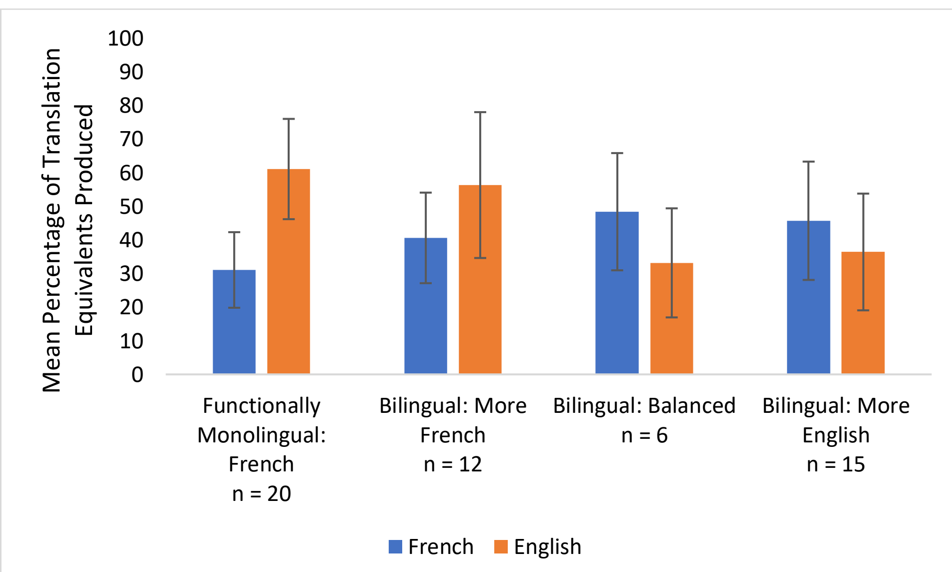
Note: n = 78

Translation Equivalents

The use of the same words in each language, translation equivalents, was explored. This was first done with the participants in Montreal who completed the tasks in English and French. Two scores were computed: percentage overlap of translation equivalents out of total French productions and percentage overlap of translation equivalents out of total English productions. The main interest was in whether the number of translation equivalents varies as function of exposure to each language. To investigate this, participants in Montreal were split into four groups based on exposure to each language based on parent questionnaires, as described earlier: Functionally Monolingual, Bilingual: More French, Balanced Bilingual, Bilingual: More English, similar to several previous studies looking at bilingual vocabulary (e.g. Elin Thordardottir, 2011, 2019). Only English-French bilinguals were included in this analysis since these were the languages for which performance was compared. Results are shown in Figure 8. The two measures that were computed were compared between French and English in each exposure group with one-tailed paired samples *t*-tests. Functionally monolingual French speakers produced a greater percentage of translation equivalents in English than in French $t(19) = 13.06$, $p < .001$, with a small effect size (Cohen's $d = .10$). Bilinguals with greater exposure to French also produced a greater percentage of translation equivalents in English than in French, $t(11) = 2.38$, $p = .036$), with a small effect size (Cohen's $d = .23$). The opposite was found in bilinguals with greater exposure to English, who produced a greater proportion of translation equivalents on the French than the English task, $t(14) = -2.33$, $p = .036$, with a small effect size (Cohen's $d = .15$). For bilinguals with balanced exposure, a non-parametric two-tailed sign test was conducted, showing these participants having greater overlap of translation equivalents on their French than English task ($p = .031$).

Figure 8

Percentage of Translation Equivalents of Total Words Produced in English or French

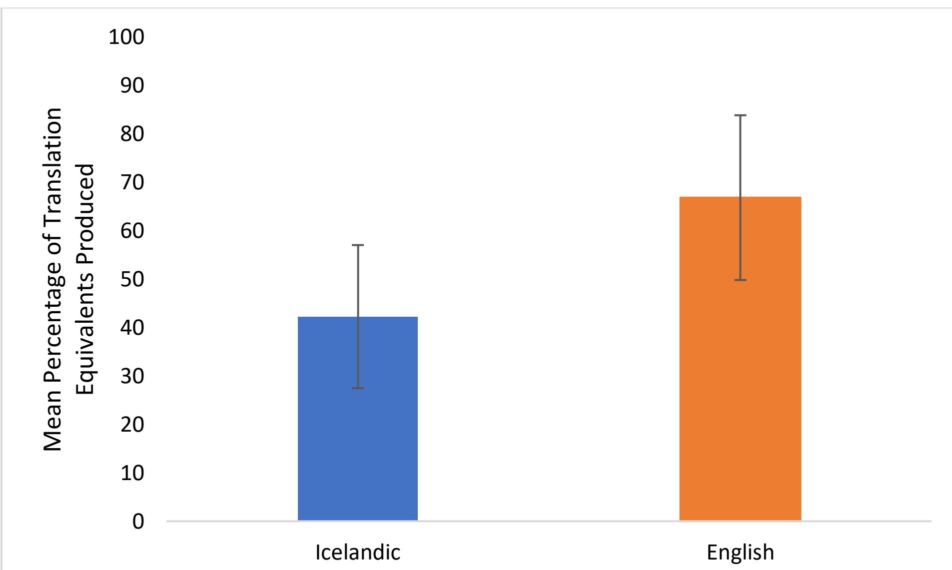


Note: Error bars denote one standard deviation.

For the L1 speakers in Iceland (those with Icelandic as their only home language) two scores were computed: translation equivalents/total Icelandic words, and translation equivalents/total English words. Results are shown in Figure 9. Just like with the Montreal participants, this provided a proportion of overlap between translation equivalents and words said in each language separately. These two values were compared with a one-tailed paired samples *t*-test. A significant difference was obtained showing greater proportion of translation equivalents produced on English than Icelandic trials: $t(31) = -7.19$, $p < .001$, with a small effect size (Cohen's $d = .19$). Participants said more unique words in Icelandic than in English.

Figure 9

Percentage of Translation Equivalents Produced by L1 Icelandic Speakers



Note: Error bars denote one standard deviation; $n = 32$.

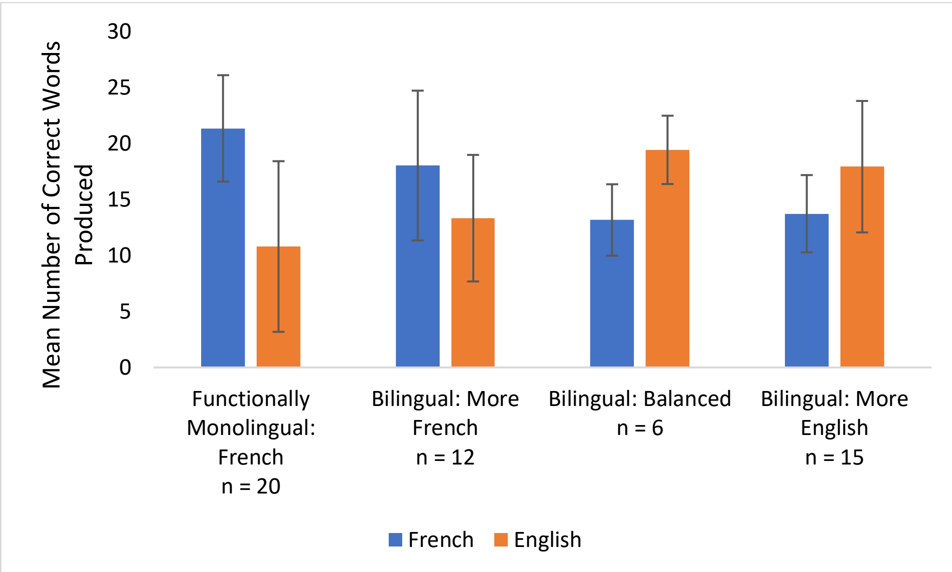
No such computations of translation equivalents were made for the Icelandic trilinguals, since it was impossible to compute their conceptual vocabulary productions including their third language. Further analysis together with production of infrequent words is planned but is beyond the scope of this thesis.

Analysis 5: Effects of Dominance

It was of interest to determine whether differences in performance might be observed between one's dominant vs. non-dominant language in participants who completed the task in more than one language. In a clinical application, it would be of great usefulness if verbal fluency scores were able to indicate which language is the stronger one for the child, since it can otherwise be difficult to identify the dominant language and because knowing which language is dominant has implications for assessment and treatment. For the Montreal participants, exposure groups were formed as was done for the analysis of translation equivalents, based on exposure to French. Mean number of words produced for each exposure group are shown in Figure 10. One-tailed paired *t*-tests were conducted for all but the balanced group, comparing performance in English and French. The functional monolinguals produced significantly more words in French than English, $t(19) = -9.23$, $p < .001$, with a very large effect size (Cohen's $d = 5.11$); the mean difference between French and English words was 10.6. Bilinguals with more French exposure also produced more words in French than English, $t(11) = -2.81$, $p = .009$, with a very large effect size (Cohen's $d = 5.85$); the mean difference between French and English words was 4.8. Bilinguals with more English exposure produced more words in English than in French, $t(14) = 2.58$, $p = .011$, with a very large effect size (Cohen's $d = 6.30$); the mean difference between French and English words was 4.2. Given the small number of balanced bilinguals, a non-parametric test, the sign test, was used to compare performance in each language. Results showed that balanced bilingual participants ($n = 6$) produced more words in English than in French, $p = .031$; the mean difference between French and English words was 6.8.

Figure 10

Total Correct Words Produced (Modified Measure) in each Language across Exposure Groups

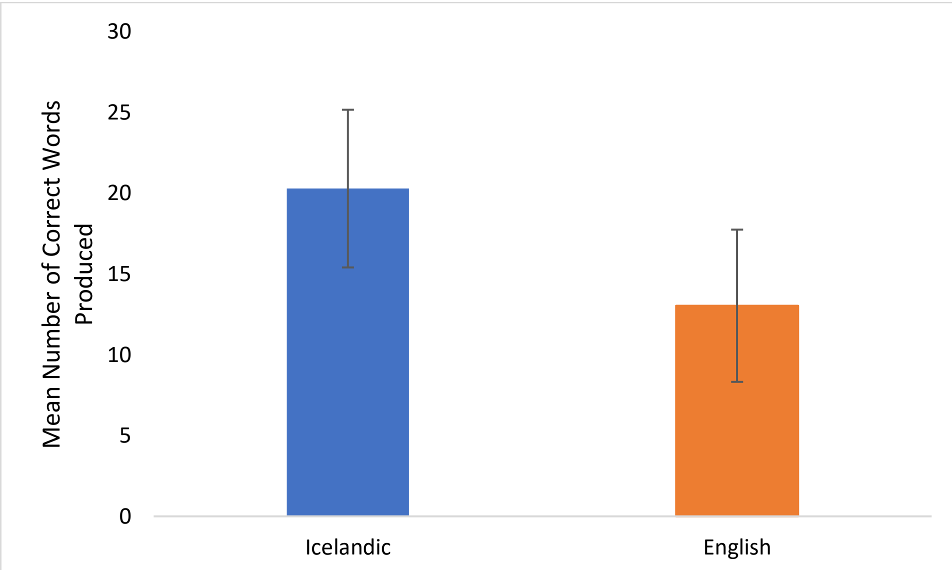


Note: Error bars denote one standard deviation.

For the Icelandic participants, analyses were first run with participants with Icelandic as their only home language (Icelandic as L1). Results are shown in Figure 11. These participants produced more Icelandic words than English words, as shown with a one-tailed paired samples *t*-test: $t(31) = 7.29$, $p < .001$, with a very large effect size (Cohen's $d = 5.62$); the mean difference between Icelandic and English words was 7.3.

Figure 11

Total Correct Words Produced (Modified Measure) in each Language for L1 Icelandic Speakers

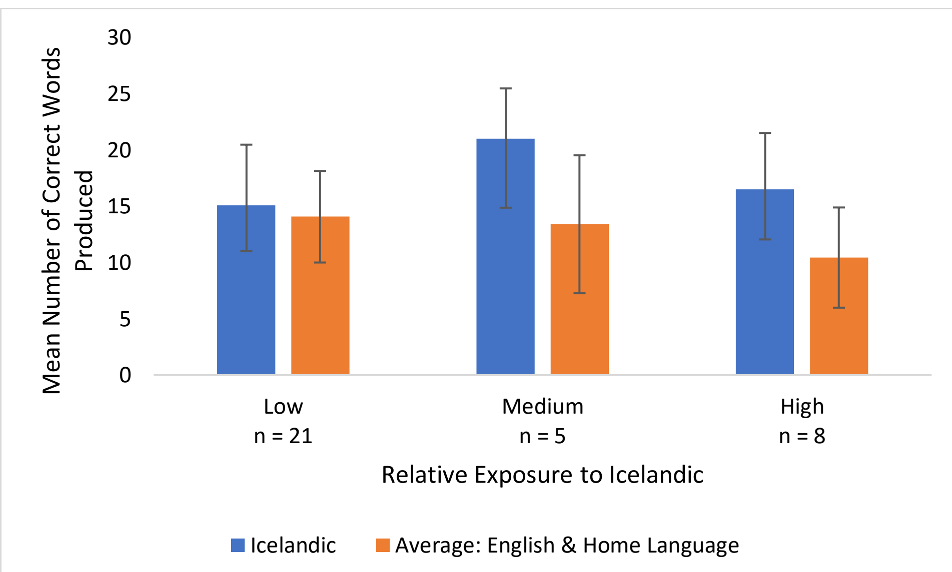


Note: Error bars denote one standard deviation; *n* = 32.

Next, this was assessed in the Icelandic trilinguals, who either had Icelandic as one of two home languages or who spoke a different language at home, in addition to having some proficiency in English. Although exposure to Icelandic was calculated from information from parent questionnaires, the relative exposure to English and the home language was not available as it is harder for parents to report on given the complex linguistic environment. Therefore, the grouping considered relative exposure to Icelandic versus other language use, be it the home language or English. Total word scores in Icelandic were compared to an averaged total words score, computed including total correct words produced in English and their total productions on the test in the home language. Exposure groups were formed based on relative exposure to Icelandic. Results are shown in Figure 12. For those with little exposure to Icelandic, the averaged English/Home Language scores were compared to Icelandic scores with a one-tailed dependent samples *t*-test. No difference was found between performance in Icelandic or averaged performance across the other two languages: $t(20) = .80$, $p = .217$. Given the small sizes of the other two groups, a non-parametric test, the sign test, was used to compare performance. Results showed no difference for those with medium exposure to Icelandic on a two-tailed sign test, $p = .375$, $n = 5$. For those with high exposure to Icelandic, a significant difference was found, on a one-tailed sign test ($p = .035$; $n = 8$); the mean difference between Icelandic and English/home language average words was 6.1. Inspection of individual scores revealed that they generally produced more words in Icelandic than the average of the two other languages (only one out of eight participants produced less words in Icelandic).

Figure 12

Total Correct Words Produced (Modified Measure) by Icelandic Trilinguals



Note: Error bars denote one standard deviation.

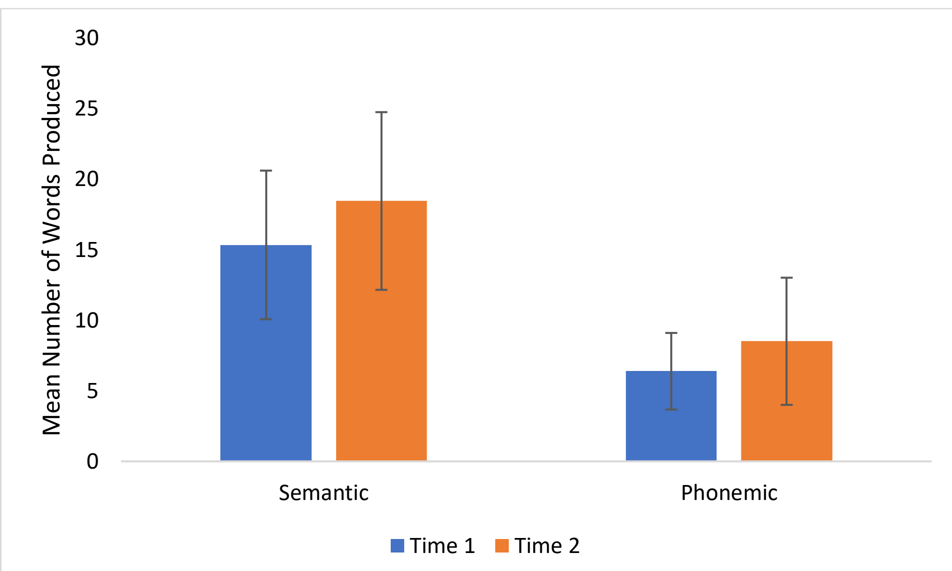
Analysis 6: Longitudinal Evolution of Verbal Fluency

A subset of participants (n = 10) participated in this project at two time points, their second participation being on average 4.69 years later (SD = 1.25). Individual performance over time was explored to evaluate the development of performance on verbal fluency tasks. One participant was excluded from this analysis for not having completed the task in the same language twice (once in French, once in English). Another participant was removed for not having completed the tasks in French, the common language. Six remaining participants were monolingual (at the time of first participation) and two were bilingual.

Participants produced more words at their second participation, particularly for semantic fluency. Participants also obtained higher vocabulary scores over time. See Figures 13 and 14.

Figure 13

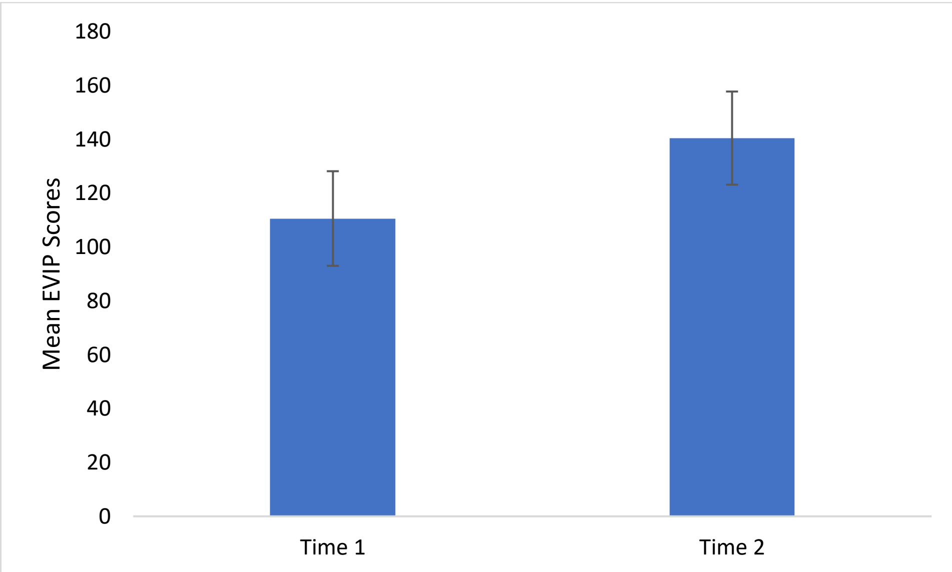
Total Words Produced (Modified Measure) on French Tasks over Time



Note: Error bars denote one standard deviation.

Figure 14

EVIP Scores over Time



Note: Error bars denote one standard deviation.

Only two of these participants completed tasks in English and French at both time points. These were observed as case studies. Participant A was first assessed at 9.1 years old and retested 4.5 years later, at age 13.6. They were a sequential bilingual, with a greater amount of exposure to English. Participant B was first assessed at 6.6 years old and was retested 6.1 years later, at age 12.7. They were a simultaneous bilingual, who was originally a balanced bilingual and at the second test date had a greater amount of exposure to French. See Figure 15 and 16 for their scores over time.

Figure 15

Participant A: Total Correct Words Produced and Receptive Vocabulary Scores over Time

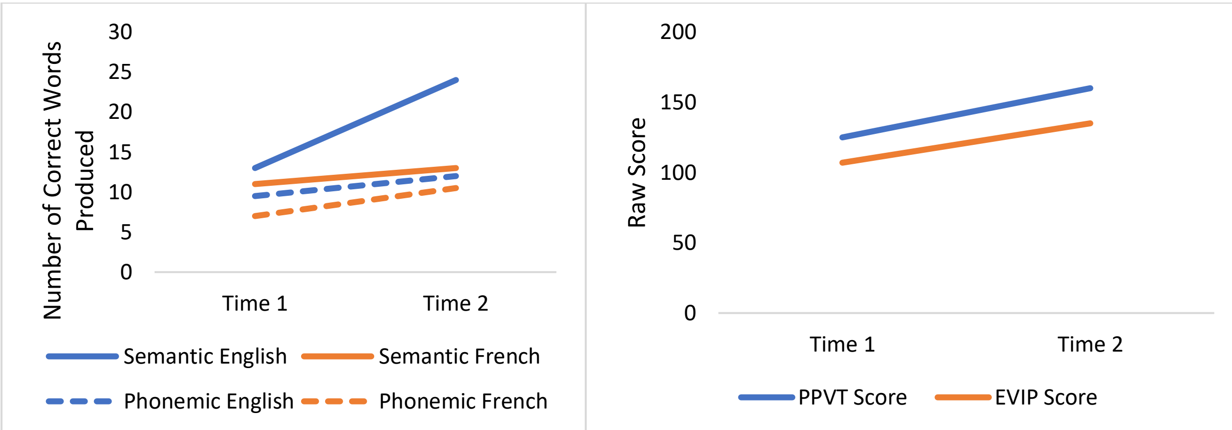
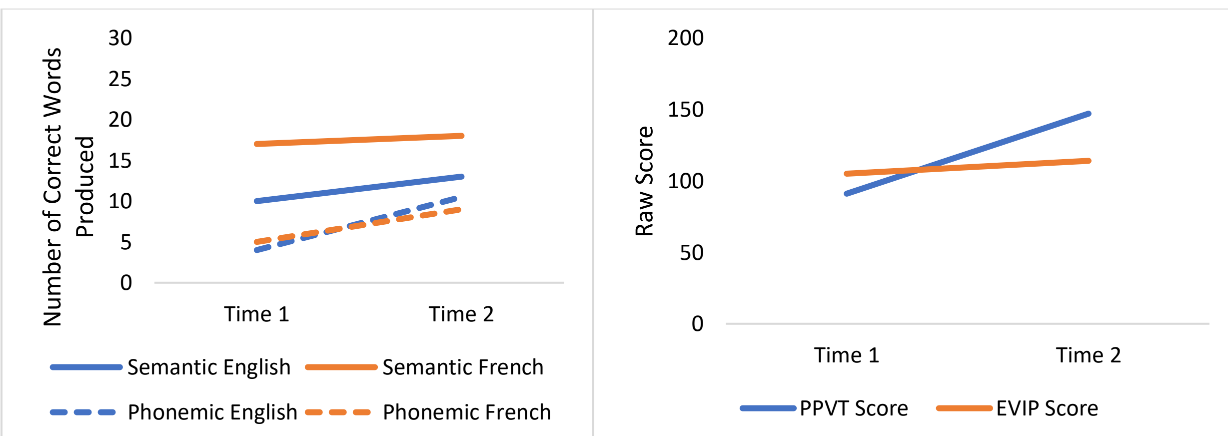


Figure 16

Participant B: Total Correct Words Produced and Receptive Vocabulary Scores over Time



In general, these participants at least produced one more word on each task over time, with no particular pattern observed between semantic vs. phonemic fluency. However, both participants had a greater change of receptive vocabulary scores in English than in French.

Discussion

The main purpose of this thesis was to assess the value of the verbal fluency task as a clinical estimate of vocabulary knowledge. The main finding was that counting the number of correct words produced on verbal fluency tasks appeared to provide several measures that are of potential use clinically. First, total correct words estimated vocabulary knowledge in English, French, and Icelandic. Both the traditional way of scoring this measure and a less-stringent, modified way were included, showing similar predictive abilities. These findings were extended to simply counting total productions in a language unknown to the experimenter, without following word-specific scoring rules, which successfully predicted self-rated comprehension and production proficiency. Further, differences in performance across languages were found to confirm predicted language dominance. Use of translation equivalents was explored and showed that participants generally produced more unique words in their dominant language. Measures of strategy use, such as mean cluster size and number of switches, were not found to be consistently useful estimators of vocabulary knowledge specifically. Taken together, these findings suggest that speech-language pathologists can count productions on verbal fluency tasks without following strict coding rules to estimate vocabulary knowledge in different languages, even when the clinician does not know this language, and that measures in two languages can confirm language dominance. This is the first time this has been demonstrated and is a valuable addition to language evaluation tools.

Estimating Vocabulary Size

Total Correct Words

The first aim of the thesis was to determine whether verbal fluency measures could predict measured vocabulary size for participants in Montreal tested in English and French. On tests of both semantic and phonemic fluency, the two measures that consistently and most strongly predicted vocabulary scores were both the traditional and modified measures of total correct words produced. This was first assessed with a task of semantic (animal) fluency.

For semantic fluency in English, total correct words significantly predicted scores on all English vocabulary tests and no French tests. In French, these measures predicted performance on both French vocabulary tests, but also on the English receptive vocabulary test, discussed below. Next, we assessed this question with data collected in Iceland. Main results were replicated with both measures of total correct words (traditional and modified) being the best predictors of receptive and expressive Icelandic vocabulary scores. These results collected in two different environments converge on the finding that measures of total correct words produced on semantic fluency tests can be used as predictors of vocabulary size. In all three languages, minimal differences in the standardized beta coefficients between the traditional and the modified scores were observed. See Tables 1 to 5 for values of the coefficients. This was interpreted as both measures being similarly useful for the purpose of predicting vocabulary size. Overall, measures of total correct words produced on a semantic fluency task appear to be strong estimators of vocabulary size.

We also assessed these same questions using measures from tests of phonemic (F, S) fluency. Again, the two measures of total correct words produced were the best and most consistent predictors. Similar standardized beta coefficients were obtained for the traditional and modified measures. English measures of total correct words predicted all English vocabulary scores as well as the French receptive vocabulary score. Parallel findings were observed in French. Language-specific predictions were made for both receptive and expressive vocabulary scores, but this measure also predicted English receptive vocabulary, discussed below. In

Icelandic, these measures both predicted receptive and expressive Icelandic vocabulary. Again, minimal differences between traditional and modified measures were obtained, and this was interpreted as the measures having similar predictive abilities. See Table 6 to 10 for values of the coefficients.

To summarize, across types of verbal fluency, languages, linguistic contexts, and in a sample containing participants of various ages, measures of total correct words accurately estimated measured vocabulary in that same language. This is the first time this has been shown, and it is of great clinical value. Verbal fluency tasks are quick (60 seconds) and easy to administer. They require few materials, including paper, pencil, and a timer. A recording device can also be used to note specific words produced, discussed in further detail below. Although specific scoring criteria have been used in the past (e.g., no superordinate category items, no sex/age variations), our findings do not support the need for such scoring to be used when the main purpose is to estimate vocabulary size, as shown by similar findings whether the traditional or modified scoring method was used. Clinicians can simply score an item as correct as long as it adheres to the category criterion (e.g., animals). Although this task does not replace a standardized test, it can be useful when such a test is not available, for example when assessing French expressive vocabulary in Quebec in a client above the age of nine. Given our findings, we consider verbal fluency to be a valuable tool for speech-language pathologists to estimate vocabulary knowledge.

As discussed, the relative links between semantic vs. phonemic fluency and vocabulary have been debated in the past (e.g., Bialystok, 2009; Paap et al., 2017; Patra et al., 2020; Shae et al., 2014; Whiteside et al., 2020). Both were judged to be important to include in the search for a useful tool to estimate vocabulary size. Both measures of total number of correct words produced on both semantic and phonemic fluency tasks successfully predicted language-specific expressive vocabulary.

Cross-language Findings

As mentioned, some cross-language predictions were also made. Some of these were consistent across languages and may be explained by factors other than language-specific vocabulary knowledge. Others were inconsistently observed and may not be of value for the current purposes. For the semantic task, French measures of total words were found to predict English receptive vocabulary. It may be that that certain factors beyond explicit word knowledge are involved. In this case, one possible explanation is that characteristics of the two languages being assessed may have played a role. Many words considered to be more advanced in English come from French, and knowing French may facilitate performance on English vocabulary tests at higher levels. In the PPVT, many of the more advanced items share cognates with French. Given this knowledge, it may facilitate selecting the proper item on the PPVT through deduction, even if participants do not necessarily know the words to be produced on an expressive test. For the phonemic fluency task, in both English and French, verbal fluency scores made cross-language predictions of receptive vocabulary scores. One possible interpretation is that these participants have high knowledge of the phonological systems of their languages, which helped them produce words on phonemic fluency tasks and led to a significant cross-language prediction. Indeed, bilingual speakers of varying exposure levels have been found to be good at non-word repetition tasks, given some minimal exposure is obtained, and performance is independent of vocabulary knowledge, even in a second language (e.g., Elin Thordardottir, 2020; Elin Thordardottir & Brandeker, 2013; Elin Thordardottir & Juliusdottir, 2013; Lee et al., 2013). Therefore, this may be a less true measure of vocabulary knowledge, rather than a measure that was inflated by phonological knowledge and possibly by knowledge of derivational and inflectional morphology. Coming back to the debate mentioned earlier, using measures of total correct words, it seems that semantic rather than phonemic fluency is more closely related to vocabulary knowledge, supporting previous findings (e.g., Bialystok, 2009).

Clusters and Switches

In this thesis, the main focus was on which words were used rather than which strategies, such as clustering and switching, were used to find them. However, we also wanted to evaluate whether search strategies provided important information on vocabulary size. Overall, mean cluster size was not found to be a useful measure of vocabulary size. It made few significant predictions, and when it did, these predictions were weak and showed no consistent pattern. Number of switches made several significant predictions. Some of these were quite strong, although not as strong as the ones made by our two measures of total correct words. Number of switches consistently predicted receptive vocabulary scores, regardless of the language of testing. It can be speculated that the ability to switch more might predict one's ability to consider all options on a receptive test, switching one's attention from one image to the next, eliminating incorrect selections, and choosing the appropriate item. Therefore, this would not be a language-specific measure, but rather a measure of one's ability to jump from one piece of information to another. Alternative explanations, possibly regarding the type of knowledge that receptive vocabulary tests assess, are likely to exist and require further investigation. Results for whether number of switches predicted expressive vocabulary were inconsistent. As shown previously, both clustering and switching are useful verbal fluency strategies, and bilinguals tend to rely on switching to produce more words (Brandeker & Elin Thordardottir, under review). Taken together with the current findings, frequent switching might be a useful strategy when large clusters are not produced, as both strategies cannot be used together, and the observed inconsistency may be due to different strategy use across participants given their varied language experiences. Therefore, search strategies such as clustering and switching do not appear to be the most valuable measures to collect with children and adolescents for the purpose of predicting vocabulary size, especially given individual differences in proficiency and language background. Further, given the wide age range in the current study, it is also possible that strategy use may be based on other skills and experiences, and that total words might more accurately reflect vocabulary knowledge regardless of such age-related factors beyond those strictly involved in vocabulary growth. In addition to these findings, the idea that these measures can be more difficult and time-consuming to score supports not using them in the clinic. We propose using our more lenient measure of total correct words (termed modified scoring in our analyses) to obtain a quick estimate of vocabulary size.

Counting Productions in an Unknown Language

Beyond needing a tool when a standardized test is unavailable, speech-language pathologists also often require ways to estimate their client's proficiency in a language they cannot assess due to lack of knowledge of this language. We assessed whether performance on a semantic fluency task could be used to predict proficiency in a language unknown to the experimenter. Number of productions in a language unknown to the experimenter successfully predicted participant self-ratings of both comprehension and production abilities in that language. A previous study on another sample of participants had shown that the same self-rating measure was highly correlated with vocabulary scores in a language-specific way, thus supporting their validity (Elin Thordardottir, 2021). This is the first study to demonstrate this use of verbal fluency. This finding suggests that when speech-language pathologists assess a client and do not know one or more of the languages that they speak, they can simply ask them to name as many animals as possible in one minute to get an estimate of vocabulary proficiency. Standardized beta coefficients were comparable whether vocabulary test scores (Receptive: .70, Expressive: .65) or self-report measures (Receptive: .55, Expressive: .72) were the outcome variables. Note that this tool is not a substitute for a formal test in the given language, but since tests are often unavailable, an estimation procedure is of great value.

During this task, some participants provided translations of the words they were producing. These were not counted in the total counts. However, to avoid having participants use up task time on providing these translations, future use of this task for this purpose should include explicit instructions not to provide them. Further, even if the speech-language pathologist has some knowledge of the language and notices some repetitions, it should be noted that repetitions were included in this measure of total productions to maintain a consistent procedure. In order to maintain a consistent procedure across clients, regardless of language, all productions should be included in the total count.

Use of Infrequent Words

Percentage of infrequent words produced did not provide indication of vocabulary size, likely since this measure covaries with the total number of words produced. However, the use of many infrequent words did appear to indicate high receptive vocabulary scores. This was observed in all three languages assessed (English, French, Icelandic). Based on our findings, if a client produces 10 infrequent words, we can be fairly confident that they would obtain a high vocabulary score. One likely explanation is that knowing many words increases the likelihood of knowing less frequently used words. This guideline is useful as it can quickly signal to a clinician whether vocabulary knowledge is expected to be high. Further, the estimation of a high vocabulary size would be strengthened if the total word count is high and there are also many infrequent words. Therefore, this quick test can be quite informative. It should be noted, however, that not using such infrequent words did not necessarily imply low vocabulary scores. It is not uncommon for language evaluation indicators to work in such an asymmetric manner. For example, many language tests have very unequal sensitivity and specificity for detecting Developmental Language Disorder (DLD; e.g., Elin Thordardottir et al., 2011) such that a low score strongly indicates the presence of DLD whereas a high score is uninformative. Therefore, number of correct words produced might still be the strongest estimator of vocabulary size, but use of infrequent words can provide some other information about vocabulary knowledge, such as which types of words (frequent vs. infrequent) are known in each language, and can hint to the kind of experience with the language a client has. Although common words do contribute to the total word count, less frequent words might be informative measures of vocabulary knowledge. Further research on the use of these words is required. The use of infrequent words may be relevant in cross-language comparisons, especially if common words are more likely than infrequent words to be included in the translation equivalents produced by participants.

Conceptual Productions and Translation Equivalents

The percentage of translation equivalents produced in each language was compared in participants tested in two languages, based on their relative language exposure. In Montreal, for functionally monolingual French speakers and bilinguals with greater French exposure, translation equivalents made up a smaller proportion of total productions in French than in English. Therefore, they produced more unique words in French. For both the bilinguals with greater exposure to English and those with balanced French-English exposure, translation equivalents made up a smaller proportion of total productions in English than in French.

Therefore, they produced more unique words in English. These findings were replicated with the L1 speakers of Icelandic. These participants said more unique words in Icelandic than in English. These findings suggest that although certain items are produced in both languages, individuals may know and easily access a greater number of additional words in their dominant language (as measured by relative exposure). Even for the balanced bilinguals, this finding was in line with differences in performance between their two languages, described below. These findings further suggest that by counting translation equivalents and computing the percentage, speech-language pathologists can obtain an estimate of unique knowledge in each language that falls in line with their language dominance.

For younger children, conceptual vocabulary, and therefore the amount of overlap and uniqueness between the two languages, can be assessed with the MacArthur-Bates Developmental Communication Inventory (MCDI; Fenson, 2007), but this tool is only useable until age three. Beyond then, conceptual and unique vocabulary is difficult to assess meaningfully. With standardized vocabulary assessments, depending on which items the participant is administered, given pre-determined rules for on which item to start and stop test administration, even on translated versions of the same test, clients may not be administered the same test items. Therefore, a true conceptual vocabulary measure may not be obtained. The verbal fluency task, rather than having a cut-off based on performance, allots equal time to all, and productions are self-generated by the participant. Therefore, it might provide a better estimate of amount of overlap between the two languages, as indicated by the present results.

Taken together with the findings regarding the use of infrequent words, verbal fluency appears to be a useful indicator of vocabulary knowledge. Although previously focused on as being interesting for executive functioning strategies, verbal fluency appears to not only estimate overall vocabulary size, but it also has other characteristics that indicate that it is sensitive to various aspects of vocabulary knowledge. Participants were likely to produce a core set of animal names across languages, but whether they included less frequent ones was associated with how well they knew the language in question. Words uniquely produced in one language, which are less likely to include common words (e.g., dog, cat), were more likely to be produced in the dominant language. Therefore, performance on this task provides information about participants' experiences in each language. This quick, simple task seems to give much information about vocabulary knowledge. Further analysis of infrequent word use and translation equivalents is planned but goes beyond the scope of this thesis.

Indicating Language Dominance

Given the need to establish a client's dominant language in an assessment, although we found this likely can be hinted at based on use of infrequent words in each language, we also explored whether verbal fluency performance in two languages could indicate language dominance, as measured by the modified measure of total correct words on the semantic fluency task. In Montreal, the functionally monolingual French speakers and the bilinguals with a greater exposure to French were found to produce more French than English words. Descriptive data showed the monolinguals producing the most French words. The bilingual participants with greater exposure to English produced more English words. This suggests that performance on verbal fluency in two languages can indicate language dominance. For the balanced bilinguals, even with the use of a non-parametric test, they were found to produce more words in English. This may be related to findings showing French monolinguals in Montreal having smaller vocabulary sizes than English monolinguals, reflecting a structural difference between the languages (Elin Thordardottir, 2005). Therefore, such differences in language outcomes might also come into play in this context. In Iceland, L1 Icelandic speakers produced more words in Icelandic than in English. Icelandic trilinguals with greater exposure to Icelandic produced more words in Icelandic than their averaged number of words in their two other languages. Again, this indicates that differences in performance across languages can confirm dominance. However, no differences between languages were observed for Icelandic trilinguals with low or medium exposure to Icelandic. One possible interpretation of this finding is that in speakers of three languages, unless a strong dominance is noted in one of the languages, as was the case with the Icelandic trilinguals with high amounts of exposure to Icelandic, performance is more equally distributed across all three languages, as was observed in another sample of trilingual adolescents (Elin Thordardottir, 2021). Another possible explanation is that without being able to determine whether participants had more exposure to English or their third language, and by averaging their scores across these two languages, perhaps some differences were obscured. This test might be particularly useful for this purpose when a clear dominant language is predicted. Therefore, semantic fluency tasks can be used to confirm predicted language dominance. It is less clear whether this task is useful to confirm a lack of dominance, given the lack of difference observed in Icelandic trilinguals with medium exposure to Icelandic, which was expected, but the diverging finding of differences between languages observed in balanced bilinguals in Montreal. This test might signal a lack of dominance, but characteristics and structures of the languages being compared should also be considered. Note that sample sizes were small, and non-parametric tests were used.

This is an important use of the verbal fluency task, as parent reports may not fully describe the language exposure children receive. It can be unclear exactly how much relative language exposure children receive in daycare or in school. When parents have a prediction, this task can help confirm it. According to our data, as presented in Figures 10 to 12, producing an additional 5 words on average can indicate dominance. A lack of such a difference might indicate balanced proficiency in the languages. This appears to be another useful and simple clinical application of verbal fluency. This test should be replicated in further research.

Longitudinal Analysis

Finally, a descriptive longitudinal case analysis of participants who completed verbal fluency tasks at two time points was completed. In general, the number of correct words produced increased over time. This is in line with research showing age effects resulting in greater production of words at later ages (e.g., Brandeker & Elin Thordardottir, under review; Filippetti & Allegri, 2011; Friesen et al., 2015; Hurks et al., 2010). This was the case for both semantic and phonemic fluency, suggesting that this may be linked to greater knowledge of words over time as well as increased phonological awareness (as suggested by Brandeker & Elin Thordardottir, under review). Indeed, vocabulary scores also increased over time. The two participants who completed the tasks in both languages experienced greater increases in scores on their English receptive vocabulary tests. It may be that this greater growth in English vocabulary is attributable to the presence of English not only at home but also in popular media and on the internet, but a larger sample size is needed to confirm this finding. For one participant, this was paralleled with a greater increase of both semantic and phonemic fluency scores in English compared to French. For the other, a larger change in English semantic fluency scores than in French was observed, but a greater but similar score in French phonemic fluency than in English was noted. Again, this might suggest a closer link between semantic rather than phonemic fluency and vocabulary knowledge, but note that this was a descriptive analysis. The bottom line is that as verbal fluency performance improves, so does vocabulary size.

Strengths and Limitations

This thesis included many strengths. For one, it included a large sample size. This allowed us to compare participants exposed to different languages and to differing degrees. Further, our participants were recruited in two different language environments, which allowed us to assess replicability within this thesis. This is the first study of its kind. It contributes a new way of using the verbal fluency task with a clear clinical application.

This thesis also included certain limitations. Although a large number of participants were included overall, when certain group comparisons were made, some groups contained few members given the smaller number of bilingual participants. This was especially the case when assessing groups with equal or possibly equal exposure to each of their languages. Further, only a descriptive analysis of the longitudinal data was possible given the small sample for Analysis 6. Future studies might consider assessing these questions with larger samples.

For the participants in Montreal, Canada, language-specificity of regressions was evaluated since many participants completed verbal fluency tasks and vocabulary measures in both English and French, However, for the Icelandic participants, only Icelandic vocabulary was measured directly. In future research, it would be of interest, especially in a more complex multilingual context like Iceland, to evaluate whether such predictions are made using measured vocabulary in English or other languages. However, the findings from Analysis 3 using self-ratings in various languages support that these successful predictions would be replicated.

Finally, as mentioned, the statistical approach required multiple comparisons, which can be seen as a necessary limitation. In an exploratory study, the many statistical analyses can be justified, pending verification in further replication studies. Again, such replication was possible even within this thesis in the two linguistic contexts, therefore results are further supported and are considered to be true findings. This thesis provided the foundation for relevant analyses to be included in future studies assessing this tool, such as a planned study on the use of this tool in children with Developmental Language Disorder.

In such future studies, it will be of interest to establish more concrete guidelines about how to interpret verbal fluency performance. This should include, in addition to those presented in this thesis, how many words might be expected on the verbal fluency task at a given age and when verbal fluency performance might indicate Developmental Language Disorder, if possible.

Conclusion

Overall, this thesis demonstrates the value of verbal fluency as an estimation tool for vocabulary knowledge and language dominance. A simple method of scoring total correct words may be the best predictor of vocabulary across languages. Beyond showing that this task can be used to predict vocabulary knowledge, this thesis has shown that even without knowledge of the language in which the task is completed, the task can be used by speech-language pathologists to obtain an estimate of language proficiency. This task is quick and easy to score. It is believed that using the verbal fluency task for this purpose will be a valuable addition to the speech-language pathologist's toolkit.

References

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2018). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult*, *5*(26), 441-451. https://doi.org/10.1080/23279095.2018.1439830

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508. https://doi.org/10.1111/opo.12131

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3-11. https://doi.org/10.1017/S1366728908003477

Bialystok, E., Craik, F. I., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522-538. https://doi.org/10.1016/j.jneuroling.2007.07.001

Bokat, C. E., & Goldberg, T. E. (2003). Letter and category fluency in schizophrenic patients: a meta-analysis. *Schizophrenia Research*, *64*(1), 73-78. https://doi.org/10.1016/S0920-9964(02)00282-7

Brandeker, M., & Elin Thordardottir. (under review).

Dubé, D., & Elin Thordardottir. (2019, November). *Predicting vocabulary size with verbal fluency scores*. Talk presented at the School of Communication Sciences and Disorders Annual Research Day, in Montreal, QC.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Dunn, L. M., & Theriault-Whalen, C. (1993). *Échelle de vocabulaire en images Peabody: Adaption Française du Peabody Picture Vocabulary Test.* Toronto, ON: Psycan.

Elin Thordardottir. (2005). Early lexical and syntactic development in Quebec French and English: implications for cross-linguistic and bilingual assessment. *International Journal of Language & Communication Disorders*, *40*(3), 243-278. https://doi.org/10.1080/13682820410001729655

Elin Thordardottir. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426-445. https://doi.org/10.1177/1367006911403202

Elin Thordardottir. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, *23*(1), 236-255. https://doi.org/10.1177/1367006917722418

Elin Thordardottir. (2020). Are background variables good predictors of need for L2 assistance in school? Effects of age, L1 amount, and timing of exposure on Icelandic language and nonword repetition scores. *International Journal of Bilingual Education and Bilingualism*, *23*(4), 400-422. https://doi.org/10.1080/13670050.2017.1358695

Elin Thordardottir. (2021). Adolescent language outcomes in a complex trilingual context: When typical does not mean unproblematic. *Journal of Communication Disorders*, *89*, 106060. https://doi.org/10.1016/j.jcomdis.2020.106060

Elin Thordardottir, & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1-16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Elin Thordardottir, & Juliusdottir, A. G. (2013). Icelandic as a second language: A longitudinal study of language knowledge and processing by school-age children. *International*

Journal of Bilingual Education and Bilingualism, *16*(4), 411-435. https://doi.org/10.1080/13670050.2012.693062

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingarian, G. (2011). Sensitivity and specificity of French language measures for the identification of Primary Language Impairment at age 5. *Journal of Speech, Language and Hearing Research, 54*, 580-597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Elin Thordardottir, Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages?. *Journal of Multilingual Communication Disorders, 4*(1), 1-21. https://doi.org/10.1080/14769670500215647

Escobar, G. P., Kalashnikova, M., & Escudero, P. (2018). Vocabulary matters! The relationship between verbal fluency and measures of inhibitory control in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *170*, 177-189. https://doi.org/10.1016/j.jecp.2018.01.012

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.

Filippetti, V. A., & Allegri, R. F. (2011). Verbal fluency in Spanish-speaking children: Analysis model according to task type, clustering, and switching strategies and performance over time. *The Clinical Neuropsychologist*, *25*(3), 413-436. https://doi.org/10.1080/13854046.2011.559481

Friesen, D. C., Luo, L., Luk, G., & Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language,* *Cognition and Neuroscience*, *30*(3), 238-250. https://doi.org/10.1080/23273798.2014.918630

Frost, J. (2019). *Regression analysis: An intuitive guide for using and interpreting linear models*. Statistics by Jim.

Gonzalez-Barrero, A. M., & Nadig, A. (2017). Verbal fluency in bilingual children with Autism Spectrum Disorders. *Linguistic Approaches to Bilingualism*, *7*(3), 460-475. https://doi.org/10.1075/lab.15023.gon

Hedden, T., Lautenschlager, G., & Park, D. C. (2005). Contributions of processing ability and knowledge to verbal memory tasks across the adult life-span. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 169-190. https://doi.org/10.1080/02724980443000179

Henry, J. D., & Crawford, J. R. (2004a). Verbal fluency deficits in Parkinson's disease: a meta-analysis. *Journal of the International Neuropsychological Society*, *10*(4), 608-622. https://doi.org/10.1017/S1355617704104141

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology*, *18*(4), 621-628. https://doi.org/10.1037/0894-4105.18.4.621

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, *42*(9), 1212-1222. https://doi.org/10.1016/j.neuropsychologia.2004.02.001

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1-27. https://doi.org/10.1017/S0305000910000759

Hurks, P. P., Schrans, D., Meijs, C., Wassenberg, R., Feron, F. J. M., & Jolles, J. (2010). Developmental changes in semantic verbal fluency: Analyses of word productivity as a function of time, clustering, and switching. *Child Neuropsychology*, *16*(4), 366-387. https://doi.org/10.1080/09297041003671184

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*(4), 381-394. https://doi.org/10.1044/1058-0360(2002/042)

Kormi-Nouri, R., Moradi, A. R., Moradi, S., Akbari-Zardkhaneh, S., & Zahedian, H. (2012). The effect of bilingualism on letter and category fluency tasks in primary school children: Advantage or disadvantage?. *Bilingualism: Language and cognition*, *15*, 351-364. https://doi.org/10.1017/S1366728910000192

Lee, H. J., Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology*, *15*(2), 375-382. https://doi.org/10.3109/17549507.2012.752866

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29-41. https://doi.org/10.1016/j.cognition.2009.08.014

Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children* Vol. 2. Clevedon, UK: Multilingual Matters.

Paap, K. R., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics, 27*(1), 50-74. https://doi.org/10.1016/j.jneuroling.2013.09.002

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*(2), 89-112. https://doi.org/10.1080/20445911.2016.1248436

Patra, A., Bose, A., & Marinis, T. (2020). Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition*, *23*(1), 204-218. https://doi.org/10.1017/S1366728918001098

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93-120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Richards, B. J., & Malvern, D. D. (1997). *Quantifying lexical diversity in the study of language development*. University of Reading, Faculty of Education and Community Studies. As cited in Elin Thordardottir, & Ellis Weismer, S. (2001). High-frequency verbs and verb diversity in the spontaneous speech of school-age children with specific language impairment. *International Journal of Language & Communication Disorders*, *36*(2), 221-244. https://doi.org/10.1080/13682820118239

Rondal (2003). *Langage oral*. In J.A. Rondal & X. Seron (Eds.), *Trouble du langage: bases théoriques, diagnostic et rééducation* (pp. 375-411). Sprimont: Mardaga. As cited in Thordardottir, E. (2016). Long versus short language samples: A clinical procedure for French language assessment. *Canadian Journal of Speech-Language Pathology & Audiology*, *40*(3). 176-197.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*(1), 17-24. https://doi.org/10.1207/S15324826AN0701_3

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, (CELF-4).* The Psychological Corporation. San Antonio, TX.

Semel, E., Wiig, E., Secord, W. A., Boulianne, L., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales–Version pour francophones du Canada.* Toronto: The Psychological Corporation.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Statistics Canada. (2017, November 29). *Census Profile, 2016 Census: Quebec [Province] and Canada [Country]*. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&SearchText=quebec&SearchType=Begins&SearchPR=01&TABID=1&G=1&Geo1=PR&Code1=24&Geo2=PR&Code2=01&type=0&B1=Language

Taler, V., Johns, B. T., & Jones, M. N. (2019). A large-scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian longitudinal study on aging. The *Journals of Gerontology: Series B, 75*(9), e221-e230. https://doi.org/10.1093/geronb/gbz003

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167-177. https://doi.org/10.1093/arclin/14.2.167

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370-378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138-146. https://doi.org/10.1037/0894-4105.11.1.138

Williams, K. T. (1997). *Expressive vocabulary test second edition (EVT™ 2).* Pearson Publishing.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure?. *Applied Neuropsychology: Adult*, *23*(1), 29-34. https://doi.org/10.1080/23279095.2015.1004574

Appendix 1

List of Infrequent English Words

Note: Words were deemed infrequent when they were tied for the 10 least produced words in each language within the linguistic context.
Words are listed alphabetically.

aardvark
albatross
alligator
amphibian
ant
anteater
antelope
ape
baboon
bat
beaver
beluga
black bear
blowfish
blue jay
boa
boar
bovine
buffalo
bug
bull
bunny
butterfly
calf
camel
canine
cardinal
caterpillar
catfish
cayote
cheetah
chicken
chihuahua
chimpanzee
chipmunk
clownfish
cobra
cold-blooded animals
companion animals
cougar
crab
crocodile
crow
deer
dinosaur
dodo bird
donkey
dove
eagle
eel/electric eel
falcon
feline
flamingo
fly
fox
gauffer
gazelle
gecko
gecko
goat
goldfish
goose
gorilla
grasshopper
guinea pig
hamster
hare
hedgehog
hen
hippo/hippopotamus
human
hyena
iguana
insect
jaguar
kangaroo
killer whale
kitten
koala
ladybug
lamb
leopard
lobster
loon
lynx
mammal
manatee
mole
moose
mosquito
octopus
orca
owl
panda
panther
parrot
pelican
penguin
pigeon
platypus
pork
puffer fish
puma
puppy
raccoon
raven
reptile
rhino/rhinoceros
rooster
salamander
scorpion

sea otter
seagull
seahorse
seal
sheep
skunk
snail
spider
squid
squirrel
starfish
sting ray
swan
tadpole
tarantula
toad
tortoise
toucan
turkey
weasel
wild animals
wildebeest
wolf
wombat
woolly mammoth
worm

Appendix 2

List of Infrequent French Words

abeille
agneau
aigle
aigle royal
alligator
anaconda
âne
ara
autruche
babouin
belette
bélier
béluga
bison
bœuf
bourdon
brebis
buffle
calmar
caméléon
canarie
canne
capybara
carcajou
cardinal rouge
caribou
castor
cerf
cerf de virginie
chameau
chaton
chauve- souris
chenille
chevreuil
chimpanzé
chinchilla
chouette
cigale
cobra
coccinelle
colibri
condor
corbeau
corneille
coucou
cougar
couleuvre
coyote
crabe
crapaud
crevette
crustacé
cygne
dinde
dindon
dingo
dinosaure
dodo
dromadaire
écrevisse
élan
étoile de mer
faucon
faucon pèlerin
félin
fennec
flamant rose
fourmi
gazelle
geai
geai bleu
gerbille
grizzly
hérisson
héron
hibou
hippocampe
hirondelle
homard
homme
huitre
humain
hyène
iguane
insecte
jaguar
jument
kangourou
kiwi
koala
lama
lamantin
lémur
lièvre
limace
lionne
loutre
lynx
macaque
mammifère
mammouth
manchot
marmotte
méduse
menthe religieuse
mille pattes
moineau
moineau
morse
mouche
mouette
moufette
moustique
mule
narval
ocelot
oïe
okapi
orang-outan
orignal
ornithorynque
orque
otarie
ouistiti
ours brun
ours noir

ours polaire
ourson
panda
panthère noire
paon
papillon
paresseux
pélican
perruche
phoque
pie
pieuvre
pigeon
pika
piranha
poisson chèvre
poisson clown
poisson rouge
poney
poulain
poulet
poux
puce
raie
raton-laveur
renne
reptile
requin blanc
rossignol
rouge-gorge
salamandre
sanglier
sangsue
saumon
sauterelle
scarabée
scorpion
souriceau
suisse
tamanoir
tamia
taupe
taureau
termite
têtard
thon
toucan
truie
truite
vautour
veau
ver de terre
vipère
volatile

Appendix 3

List of Infrequent Icelandic Words

ær
álft
alpaca
ánamaðkur
andarungi
antilópa
asni
áttfætla
bambi
bjalla
bleikja
blettatígur
broddgöltur
býfluga
dádýr
dúfa/ur
elgur
endur
engispretta
fálki
fiðrildi
fíll (bird)
fjallaljón
flamingó
folald
froskdýr
gæs
gári
gaupa
geitungar/ur
göltur
górilla/ur
grágæs
grís/grisir
gullfiskur
gylta
haförn
háhyrningur
hani
héri
himbrimi
hlébarði/ur,
höfrungar/ur
hrafn
hreindýr
hrossagaukur
humar
húsfluga
hvítháfur
hvolpar/ur
hýena/ur
íkornar/i
jagúar
járnsmiður
kakkalakki
kálfur
kalkúnn
kameldýr
kameleón
kengúra
kettlingar/ur
kjúklingur
kóala/koalabjörn
kóbra
kolkrabbi
krabbar
kría
krossfiskar/ur
krummi
labrador
lamadýr
lax
leðurblaka
letidýr
lirfa
lama
lóa
lúða
lundi
lús
manneskja/ur
margfætla
marglyttar/ur
maríubjalla
maríuhæna
maur/maurar
minkar/ur
mörgæs
músarrindill
mýfluga
naggrís
nashyrningur
naut
óðinshani
önd
ormar/ur
örn
padda/pöddur
panda/pandabjörn
pokarotta
puma
rækjur
rauðmagi
refur/refir
risaeðla
rjúpa
rostungur
sæhestur
sandkoli
síberíutígrisdýr
síld
síli
silungur
simpansi
skarkoli
skata
skógarþröstur
skordýr
skúnkur
slanga
smokkfiskar
snígill
spói
sporðdreki
steinbíturstorkur
strútur

svartbjörn
sverðhákarl
tarantúla
þorksur
þröstur
þúsundfætla

tjaldur
túnfiskur
ugla
úlfaldi
úlfur
ungi

urriði
vespa
villiköttur
villisvín
ýsa