American Journal on Intellectual and Developmental Disabilities

# Evaluating verbal fluency outcome measures in children with Down syndrome

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | AJIDD-D-21-00020R2 |
| **Article Type:** | Research Report |
| **Keywords:** | Down syndrome; Children; adolescents; verbal fluency; outcome measures |
| **Corresponding Author:** | Anna J Esbensen, PhD<br>Cincinnati Children's Hospital Medical Center<br>Cincinnati, OH UNITED STATES |
| **First Author:** | Anna J Esbensen, PhD |
| **Order of Authors:** | Anna J Esbensen, PhD |
| | Catelyn Smeyne |
| | Emily Schworer |
| | Shequanna Belizaire |
| | Emily Hoffman |
| | Dean Beebe |
| | Susan Wiley |
| **Manuscript Region of Origin:** | UNITED STATES |
| **Abstract:** | This study evaluates the psychometric properties of a verbal fluency task for potential use as an outcome measure in future clinical trials involving children with Down syndrome. Eighty-five participants attempted a modified version of the NEPSY-II Word Generation Task at two time points. In the full sample, the measure fell below a priori reliability and feasibility criteria, though feasibility of the semantic trials were higher than feasibility of the phonemic trials. Performance on the measure correlated with chronological age and IQ scores, and no sex-related effects were found. Additional analyses suggested that the semantic verbal fluency trials might be appropriate for children with Down syndrome 10 years of age and older. |

# Evaluating Verbal Fluency Outcome Measures in Children with Down Syndrome

## Abstract

This study evaluates the psychometric properties of a verbal fluency task for potential use as an outcome measure in future clinical trials involving children with Down syndrome. Eighty-five participants attempted a modified version of the NEPSY-II Word Generation Task at two time points. In the full sample, the measure fell below *a priori* reliability and feasibility criteria, though feasibility of the semantic trials were higher than feasibility of the phonemic trials. Performance on the measure correlated with chronological age and IQ scores, and no sex-related effects were found. Additional analyses suggested that the semantic verbal fluency trials might be appropriate for children with Down syndrome 10 years of age and older.

*Keywords:* Down syndrome, children, adolescents, verbal fluency, outcome measures

**Evaluating verbal fluency outcome measures in children with Down syndrome**

Down syndrome (DS) is one of the most common causes of intellectual disability, and it is typically associated with clinical challenges with executive functioning and language skills (Abbeduto et al., 2007; Fidler et al., 2020; Parker et al., 2010). Individuals with DS present with challenges in verbal short-term memory and shifting tasks that are further complicated by critical deficits in expressive language (Fabbretti et al., 1997; Fidler, 2005; Fidler et al., 2020). These impairments in executive control and language highlight a need for effective therapeutic interventions to support cognitive and behavioral outcomes in individuals with DS (Best et al., 2011; Daunhauer et al., 2014; Fidler et al., 2020; Naess et al., 2011; Sabat et al., 2020).

Meaningful clinical findings remain limited in pharmacological and behavioral trials pursuing interventions targeting executive function and language in DS (Liogier d'Ardhuy et al., 2015; de la Torre et al., 2016; Hart et al., 2017; Hewitt et al., 2005; Kishani et al. 2010; Keeling et al., 2017; Wright et al., 2013). These null results may be in part due to a lack of psychometrically evaluated outcome measures appropriate for individuals with DS and other populations with intellectual disability (de la Torre et al., 2016; Edgin et al., 2017; Esbensen et al., 2017; Hart et al., 2016; Heller et al., 2006; Keeling et al., 2016). Preliminary work has identified parent-report measures of executive function for use in children and adolescents with DS (Esbensen et al. 2019), but the field still lacks well-characterized in-person neuropsychological assessments of executive function and language constructs within this population (Esbensen et al., 2017; Eunice Kennedy Shriver National Institute of Child Health and Human Development, 2015). In particular, outcome measures for clinical trials designed for children and adolescents with DS must (a) minimize floor effects, to better measure change and to identify group heterogeneity; (b) have known within-individual sensitivity to change, and

within-sample sensitivity, such as to age, sex, and IQ, to guide study designs and covariate; and (c) have adequate test-retest reliability, to measure potential change across time (Edgin et al., 2010; Esbensen et al., 2017). Without evidence for these characteristics, clinical trials in pediatric DS could fail because weak outcome measures obscure intervention effects.

Verbal fluency tasks are some of the most common neuropsychological measures of executive control and language production, and their use as outcome measures in clinical trials in adolescents and adults with DS suggest they may be useful in younger populations as well. The semantic fluency subtest of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), a measure appropriate for individuals 12-89 years, demonstrates limited floor effects and sensitivity to language and cognitive ability in adolescents and young adults with DS (Liogier d'Ardhuy et al., 2015). Semantic verbal fluency tasks, such as the subtest of the RBANS for which test-takers are asked to quickly generate words that fall within given categories (e.g., foods, animals), assess semantic memory, categorization, filtering/inhibition, and retrieval skills (Randolph et al., 1998). Though not extensively evaluated in individuals with DS, verbal fluency tasks can also measure phonemic fluency – the ability to identify the appropriate sounds utilizing both semantic and phonemic memory stores. Phonemic verbal fluency tasks are similar to the semantic tasks but require words that start with a given sound or letter instead of words within a given semantic category. These tasks require more exhaustive search of the lexicon, resulting in a greater demand for executive control than the semantic counterpart (Costafreda et al., 2006; Hurks, 2006; Nash & Snowling, 2008; Troyer et al., 1997; Troyer et al., 1998; Troyer, 2000). Additionally, phonemic verbal fluency tasks require adequate articulation and phonological skills to support initial phoneme differentiation of spoken words (Marcell, 1995; Snowling et al., 2002; Nash & Snowling, 2008). Beyond being important indicators of language-based executive

functioning, verbal fluency tasks reflect conversational skills that are often important to families in anecdotal reports (Del Hoyo et al., 2015).

Verbal fluency tasks have been used in typically developing (TD) children ages 2-18 years and in children with intellectual disabilities ages 9-19 years (Kemp et al., 2001; Memisevic et al., 2017; Nash & Snowling, 2008; Pennington et al., 2003). Overall, verbal fluency productivity increases with age and receptive vocabulary in TD children and youth with DS (Liogier d'Ardhuy et al., 2015; Troyer, 2000; Riva et al., 2000). However, there is no significant relationship between phonemic verbal fluency and receptive vocabulary in DS (Nash & Snowling, 2008). Children with DS also perform worse overall on both semantic and phonemic tasks compared to TD children, even when matched on receptive vocabulary, which suggests that vocabulary is required but not sufficient for children with DS to perform comparably to TD peers (Nash & Snowling, 2008). Additionally, TD children show slight sex differences, with female participants typically outperforming their male counterparts (Harrison et al., 2000; Troyer, 2000; Riva et al., 2000). Previous research has shown no sex-related effects on semantic verbal fluency performance in individuals with DS, but no studies to date have evaluated sex-related effects on the phonemic verbal fluency task when completed by individuals with DS (Del Hoyo et al., 2015).

A few studies have described multiple variations of the verbal fluency task in individuals with intellectual disability and DS and suggest promise for their potential utility in future clinical trials (Ball et al., 2008; Azuma, 2004; Del Hoyo et al., 2015; Nash & Snowling, 2008; Pennington et al., 2003). The semantic verbal fluency task has minimal floor effects when completed by individuals with DS, though the phonemic task has not been directly evaluated in this group (Ball et al., 2008; Esbensen et al., 2017). Similarly, test-retest reliability has been

established only for semantic (but not phonemic verbal fluency) among adolescents and adults (not younger children) with DS (Lioger d'Ardhuy et al., 2015); yet, previous research has documented adequate test-retest reliability of the full range of verbal fluency tasks within TD school-aged populations (Harrison et al., 2000; Snow et al., 1988). Additionally, relevant research has been limited to the primary outcomes of total unique words produced, ignoring verbal fluency errors (incorrect or unrelated words) and perseverations (repetitions of correct words). These error scores may provide complementary information, given the commonality of floor effects on measures completed by children with DS (Edgin et al., 2010). Thus, while there is promising evidence that verbal fluency measures may be useful tools for measuring executive function and language in future clinical trials involving children with DS, there are gaps in our scientific knowledge of their psychometric properties in this population.

**Current Study**

The current study seeks to expand the list of potential outcome measures for use in clinical trials in DS by examining the feasibility and psychometrics of a modified version of the NEPSY-II Word Generation Task for use with children and adolescents with DS. The first aim of the study is to determine the feasibility of the verbal fluency task by describing the participants who are unable to complete the semantic and phonemic variations of the task, able to do so but receive "floor" (lowest possible) scores when compared to published norms, or able to do so with above-floor scores. Due to the known challenges of phonological and executive functioning in DS, we predict higher feasibility on the semantic verbal fluency trials than on the phonemic verbal fluency trials (Charchat-Fichman et al., 2011; Nash & Snowling, 2008; Laws, 2004). The second aim of the study is to describe verbal fluency task performance distributions and the correlations of performance with broader developmental domains (age, sex, IQ). Based on

previous findings in individuals with DS, we expect participant age and IQ scores, but not sex, to be correlated with performance on the verbal fluency task (Liogier d'Ardhuy et al., 2015; Del Hoyo et al., 2015; Nash & Snowling, 2008; Pennington et al., 2003). The third aim of the study is to evaluate test-retest reliability and practice effects on the verbal fluency task for children and adolescents with DS. We expect the semantic and the phonemic trials to show adequate test-retest reliability and negligible practice effects when completed by children and adolescents with DS because adequate reliability has been observed in TD populations and adolescents with DS on similar variations of the verbal fluency task (Liogier d'Ardhuy et al., 2015; Harrison et al., 2000; Snow et al., 1988). Finally, we aimed to establish clinical guidelines for the administration of the verbal fluency task for children and adolescents with DS based on observed performance and psychometric evaluation.

## METHOD

### Participants

Eighty-five children and adolescents with DS, between the ages 6-19 years old (6-9 years old = 21.18%, 10-19 years old = 78.82%, *M* = 12.64 years, *SD* = 3.49), participated in the current study as part of a community-based study evaluating cognitive outcome measures in school-age children with DS. A portion of the participants are also included in publications related to executive functioning (Schworer et al. 2021; Will et al., 2021). The age range used in this study was consistent with clinical trials highlighting the need for more sensitive outcome measures for children and adolescents with DS ages 10 years and over (Kishani et al., 2010). Children between the ages of 6 – 10 years were included to prepare for future clinical trials that will include younger children with DS.

Participants were primarily non-Hispanic (94.1%) and Caucasian (88.2%), with representation from African-American (3.5%), Asian (4.7%) and mixed race (3.5%) participants. Fifty-six percent of participants were male, and standard full-scale IQ scores on the Kaufman Brief Intelligence Test, Second Edition ranged from 40 to 73 ($M$ = 44.73, $SD$ = 7.92; Kaufman & Kaufman, 2004).

**Procedure**

Participants were recruited from a pediatric medical center, regional DS clinics, and flyers distributed by local DS associations. The study protocol was approved by the institutional review board at the medical institution via the Streamlined, Multisite, Accelerated Resources for Trials (SMART) IRB platform. Participants eligible for the study had a documented diagnosis of DS and spoke English as their primary language. Additionally, children were screened for a nonverbal mental age of 36 months per parent verbal report to estimate the child's ability to attempt a portion of the measures in the broader study, though no participants were excluded for having a mental age below 36 months alone. Children and adolescents were excluded from the study if they had sensory impairments (i.e., deafness or blindness) that would have interfered with a valid administration of the study measures.

Eligible participants were evaluated twice, two weeks apart. At Time 1, participants completed an IQ assessment. The verbal fluency task was completed during a battery of neuropsychological assessments at Time 1 and was repeated at a retest visit two weeks later (Time 2).

**Measures**

**Intelligence.** *Kaufman Brief Intelligence Test – 2nd Edition (KBIT-2; Kaufman & Kaufman, 2004).* The KBIT-2 is an assessment of cognitive ability used with children and adults

and suggested for studies in DS as an appropriate measure of overall intelligence (Edgin et al., 2010; Kaufman & Kaufman, 2004). Individual KBIT-2 Verbal and Nonverbal raw scores were used to assess the association between the verbal fluency trials to avoid floor effects common with the KBIT-2 full scale IQ score (Edgin et al., 2010; Hamburg et al., 2019).

**Verbal fluency.** The current study used a modified version of the Word Generation Task from the Neuropsychological Assessment of Children, 2nd Edition (NEPSY-II) test battery. The NEPSY-II has been normed on 1200 TD children between 3 and 16 years of age and has been used to successfully assess the cognition of adults with DS (Esbensen et al., 2017; Korkman et al., 2007). The Word Generation Task includes two semantic trials (animal fluency, food/drink fluency) and two phonemic trials (/S/ fluency, and /F/ fluency; Korkman et al., 2007). The NEPSY-II Word Generation semantic trial utilizes similar instructions to the semantic fluency task of the RBANS; yet is age appropriate for younger children. For example, the RBANS prompts for generation of fruits or vegetables, whereas the NEPSY-II Word Generation Task prompts for broader categories (e.g., food) (Korkman et al., 2007; Randolph et al., 1998). In line with previous research involving children with DS, the current study modified the phonemic trials of the Word Generation Task by using the phonemes /B/ and /T/ instead of the phonemes /S/ and /F/ (Nash & Snowling, 2008). This modification stems from research suggesting that stop consonants are one of the first to develop in TD children and in children with DS (Kumin et al., 1994; Nash & Snowling, 2008).

For each trial, the child was instructed to generate as many responses as they could in a 60-second time period. Participants were verbally given the semantic category or letter/sound and two example responses were provided by the examiner before beginning the trial. Examiners tracked the number of correct responses, the number of intrusions (i.e., words that did not fit the

category), and the number of perseverations (i.e., repetitions of a correct response). If intrusion responses were repeated, each instance was counted as a single intrusion (Korkman et al., 2007).

Administration of the verbal fluency tasks conformed closely to the test manual, but scoring differed slightly. Specifically, to differentiate if children were only repeating sample items versus generating novel responses, and to facilitate scoring perseverations, single repetitions of the examples provided by the examiner during the task instructions (cat/dog, pizza/milk, box/bus, tree/top) were counted as correct responses on all trials, with only subsequent repetitions counted as perseverations. Additionally, given the frequency of correct responses with proper nouns, they were counted as correct responses on the /B/ and /T/ trials, though these responses are categorized as intrusions on phonemic verbal fluency trials in standard administration (Korkman et al., 2007). These modifications were implemented to avoid the rejection of meaningful responses provided by the participants. In doing so, we aimed to maximize the range of performance across participants and to minimize floor effects common in assessment of children with DS (Esbensen et al., 2017).

**Analysis plan**

The first aim of the study was to describe the feasibility of the verbal fluency tasks (semantic and phonemic variations). Feasibility was defined for each verbal fluency trial (animal, food, /B/, /T/) as the number and corresponding percentage of participants who were able to provide one or more correct responses at Time 1 and Time 2. Consistent with previous research evaluating outcome measures for children with ID, the feasibility threshold used in the current study was > 80% (Hessl et al., 2016). If applicable, missing data categories were used to describe the reasons for non-completion.

The second aim of the study was to examine performance distributions on the semantic and phonemic verbal fluency tasks, and to evaluate the association between verbal fluency performance and broader developmental domains (age, sex, IQ). We limited analyses to participants who completed the tasks at both time points, to support the development of appropriate outcome measures for future clinical trials in DS that will require repeated assessment visits (Edgin et al., 2010). Descriptive statistics (mean, median, minimum, maximum) were calculated for each variable, including levels of skewness and kurtosis to evaluate the normality of score distributions. A skew statistic between -1.00 and 1.00 and a kurtosis statistic between -3.0 and 3.0 were considered acceptable. Nonparametric statistical analyses were used if skew or kurtosis were observed. Floor effects were computed using the percentage of participants who were either unable to complete the task or who received the lowest possible raw score, with floor effects ≤ 20% considered acceptable. In addition, we accounted for the repetition of examiner examples during task instructions by calculating the frequency by which they occurred for all tasks, including the number of participants who listed the examples as their only correct responses. We also examined the inclusion of proper nouns as correct responses during the phonemic trials to determine whether the addition of proper nouns affected productivity on the phonemic trials. Bivariate Pearson and Spearman correlations were used to examine the relationship between age and IQ on the number of correct items, the number of intrusions, and the number of perseverations produced during the task. Pearson correlations were used for approximately normally distributed data and Spearman coefficients were used when skewness and kurtosis were observed. Correlations between age, IQ and verbal fluency performance outcomes were categorized as weak (< .30), moderate (.30 - .70), or strong (> .70;

Akoglu, 2018). Sex differences in performance on fluency tasks were evaluated using *t*-tests for normally distributed data and Mann-Whitney U tests for nonparametric data.

The third aim of the study was to describe two-week test-retest reliability and practice effects on the semantic and phonemic verbal fluency tasks. Test-retest reliability was analyzed with intraclass correlation (ICC) and was categorized as poor (< .50), moderate (.50 - .74), good (.75 - .90), or excellent (> .90; Koo & Li, 2016). Practice effects were examined using paired samples *t*-tests for performance at Time 1 and Time 2, with Cohen's $d < .20$ categorized as negligible practice effects (Cohen, 1988).

To support administration of verbal fluency tasks for future clinical trials in DS, we aimed to establish appropriate inclusion criteria recommendations for studies that plan to use these measures. To address this aim, we described the likelihood of verbal fluency task completion based on chronological age and IQ scores using sensitivity and specificity probabilities. As applied here, sensitivity uses the performance of the participants in the sample to determine the probability that persons within given benchmarks are able to complete the task. Sensitivity is calculated as the proportion of children that were correctly identified as being able to complete the verbal fluency task given the benchmark cut-offs set by age and IQ. Specificity uses a similar calculation to determine the probability that participants within given benchmarks are unable to complete the task. Specificity is calculated as the proportion of children that were correctly identified as *not* being able to complete the verbal fluency task given the benchmark cut-offs set by age and IQ. Examined benchmarks included age cut-offs of 10 years and older with no IQ restriction, IQ > 40, IQ > 45, and IQ > 50. Based on sensitivity and specificity probabilities and the age ranges of previous clinical trials in DS (i.e., Kishani et al., 2010), we conducted feasibility and reliability analyses on an age-restricted sample to evaluate whether the

verbal fluency tasks were psychometrically sound when completed by participants with DS 10 years of age and older.

## RESULTS

### Aim 1: Examiner administration and feasibility of the verbal fluency tasks

**Feasibility.** Feasibility of the verbal fluency task is presented in Table 1. Overall, the verbal fluency task failed to reach the feasibility threshold (> 80%), but was higher for the semantic trials (74.1-75.3%) than for the phonemic trials (58.8-60.0%). For the two semantic trials (animal/food), one participant (1.2%) completed the animal fluency trial but verbally refused to complete the food fluency trial. Of the remaining 84 participants who were presented with the semantic trials, 14.3% ($n$ = 12) failed to complete the task due to limited verbal abilities, such as apraxia, language delays, and selective mutism; 2.4% ($n$ = 2) failed to complete the task due to a lack of understanding task instructions (i.e., participants had verbal abilities but did not provide any response); and 8.3% ($n$ = 7) failed to complete the task due to behavioral noncompliance such as avoidant or disruptive behaviors. For the phonemic trials, one participant (1.2%) completed the /B/ fluency trial but verbally refused to complete the /T/ fluency trial. Of the remaining participants who attempted the phonemic trials, 19.0% ($n$ = 16) failed to complete the task due to limited verbal abilities, 14.3% ($n$ = 12) failed to complete the task due to a lack of understanding task instructions, and 8.3% ($n$ = 7) failed to complete the task due to behavioral noncompliance.

**Intrusion and perseveration errors.** The number of participants who produced intrusions and perseverations at Time 1 are presented in Table 1. Intrusions occurred in 34.1% of participants on the semantic animal trial, 26.9% on the semantic food trial, 60.8% on the phonemic /B/ trial, and 60.0% on the phonemic /T/ trial. About one-half of participants

demonstrated perseverations, with 55.3% demonstrating perseverations on the semantic animal trial, 50.8% on the semantic food trial, 51.0% on the phonemic /B/ trial, and 50.0% on the phonemic /T/ trial.

**Aim 2: Performance distributions on the verbal fluency tasks**

Distributions of verbal fluency task performance are presented in Table 2. There were no skewness or kurtosis concerns for the number of correct items produced on the verbal fluency tasks (See Table 2). Floor effects were observed on all fluency trials (> 20%), and included participants who did not complete the task and participants who scored zero. Twenty-four percent of participants demonstrated floor effects on the semantic animal trial, 25.9% on the semantic food trial, 40.0% on the phonemic /B/ trial, and 41.2% on the phonemic /T/ trial. One participant (1.8%) completed the phonemic /T/ verbal fluency trial but received the lowest score on the measure (zero correct responses). This participant produced intrusions during the allotted task administration period. No participants received the lowest score on the semantic (animal and food) or the phonemic /B/ fluency trials.

To address the inclusion of examiner examples in the total number of correct responses, frequencies of the use of the examples were summed. Over half of participants repeated the examples provided during the semantic animal (63.5%), the semantic food (56.5%), the phonemic /B/ (56.5%), and the phonemic /T/ (50.6%) task instructions, but few participants used the examples as their only correct responses. On the semantic animal and semantic food verbal fluency trials, no participants used the prompts provided (Cat/Dog, Pizza/Milk) as their only correct responses. On the phonemic /B/ fluency trial, 13.7% ($n$ = 7) of participants used the prompts provided (Bus/Box) as their only correct response(s). On the phonemic /T/ fluency trial,

10.0% ($n$ = 5) of participants used the prompts provided (Tree/Top) as their only correct response(s).

During the phonemic verbal fluency trials, 13 participants (25.5%) used proper nouns during the /B/ verbal fluency trial and 10 participants (20.0%) used proper nouns during the /T/ verbal fluency trial. The inclusion of proper nouns in the total number of correct responses increased productivity outcomes on the /B/ verbal fluency trial ($t$ = 3.79, $p$ < .001) and the /T/ verbal fluency trial ($t$ = 3.23, $p$ < .05) compared to scores when not including proper nouns, but effect sizes were small ($d$ < .20). On average, when pronouns were included, correct totals increased by less than one response on the phonemic /B/ (difference of .32) and /T/ trials (difference of .24).

Age was negatively correlated with the number of perseverations produced on the Animal Fluency Trial ($\rho$ = -.32), but was not significantly correlated with any other scores. The KBIT-2 Verbal raw scores were positively related to the number of correct responses on the semantic animal and food fluency trials (See Table 2). The KBIT-2 Nonverbal raw scores were positively related to the total number of correct responses produced on all tasks (see Table 2.). KBIT-2 Verbal and Nonverbal raw scores were not related to the intrusion or perseveration errors produced during most tasks, with the except of the positive relation between the KBIT-2 Verbal raw scores and the phonemic /T/ trial perseverations. There were no sex-related effects on verbal fluency performance (See Table 2).

**Aim 3: Reliability of the verbal fluency tasks**

**Test-retest reliability and practice effects.** The number of correct responses produced on the semantic fluency trials and on the phonemic /T/ fluency trials demonstrated moderate to good test-retest reliability (ICC = .60 - .74). The number of correct responses produced on the

phonemic /B/ fluency trial showed poor test-retest reliability (ICC < .50; See Table 3). No practice effects were identified on any verbal fluency trial between Time 1 and Time 2 ($p > .05$ and $d < .20$; See Table 3).

Intrusions on the phonemic /T/ verbal fluency trials had moderate test-retest reliability (ICC = .67), but intrusions on the remaining fluency trials demonstrated poor test-retest reliability (ICC < .50; see Table 3). Perseverations on the semantic animal verbal fluency trials had moderate test-retest reliability (ICC = .62). Perseverations on the semantic food and the both phonemic verbal fluency trials had poor test-retest reliability (ICC < .50). No practice effects were identified on error scores between Time 1 and Time 2, although Cohen's *ds* were >.20 on the semantic food and phonemic /B/ perseverations (See Table 3).

**Aim 4: Clinical guidelines for administration of the verbal fluency tasks in DS**

**Sensitivity and specificity.** Although feasibility on the verbal fluency tasks did not meet *a priori* study criterion, more than half of study participants were able to complete the semantic and the phonemic trials. This suggests that the verbal fluency tasks may be appropriate for some participants with DS, such as participants who are older or participants who possess higher cognitive abilities. Sensitivity and specificity statistics were used to evaluate the characteristics of children and adolescents with DS who can complete the verbal fluency task (i.e., chronological age and IQ; see Table 4). Further illustration of participant characteristics that correspond with those able and unable to complete the semantic and the phonemic verbal fluency trials are provided in Figures 1 and 2.

Although dividing the sample out by Standard Full-Scale IQ did not improve sensitivity and specificity above *a priori* study criteria, division by age did do so; we observed high sensitivity for the semantic (92%) and the phonemic (92%) trials in children above age 10 (see

Table 4). In this age-restricted subsample ($n$ = 63), the feasibility of the semantic verbal fluency trials met *a priori* criteria (84.1%), but the feasibility of the phonemic verbal fluency trials did not (73.0%; see Table 5). The semantic fluency trials and the phonemic /T/ verbal fluency trial had moderate test-retest reliability (ICC = .59 – .73), and the phonemic /B/ fluency trial had poor test-retest reliability (ICC < .50; see Table 6.). Though Cohen's *ds* were >.20 on the semantic food and the phonemic /B/ perseverations, all trial types had no significant practice effects between Time 1 and Time 2 (see Table 6).

## DISCUSSION

The recent rise in pharmacological and behavioral clinical trials to support cognitive and behavioral outcomes in DS highlights a need for empirically-based and psychometrically sound outcome measures to accurately assess language and executive function performance and identify potential change over time. In line with the National Institutes of Health Down syndrome working group meeting in 2015 (Esbensen et al. 2017), the overarching goal of this study was to examine the psychometric properties of a verbal fluency task for potential use in future clinical trials measuring language and executive functioning skills in children and adolescents with DS. As summarized in Table 7, our results suggest that verbal fluency tasks may be a viable measure of language and executive function for some participants, but not others. The semantic variation (in which words are generated within a specific sematic category rather than starting with a given letter or sound), may be an appropriate outcome measure for children and adolescents with DS 10 years of age and older. However, the phonemic variation remains problematic even in that age range, and neither variant performs well in a broader age range that includes younger children with DS.

### *Feasibility*

Contrary to previous findings, feasibility on the verbal fluency tasks for a broad sample of children with DS ages 6-19 years did not meet study criteria (< 80%; Nash & Snowling, 2008; Pennington et al., 2003). In our full sample, the semantic verbal fluency trials had higher feasibility (animal, 75.3%, food = 74.1%) than the phonemic verbal fluency trials (/B/ = 60.0%, /T/ = 58.8%). This result was expected given how often children with DS have compromised phonological and executive function skills that are required for completion of the phonemic verbal fluency task (Costafreda et al., 2006; Fidler et al., 2020; Laws, 2004; Marcell, 1995; Nash & Snowling, 2008). Participants who were able to complete the phonemic fluency trials were also able to complete the entire verbal fluency assessment, with the exception of one participant who verbally refused to complete the /T/ fluency task. This finding further suggests the more challenging nature of the phonemic fluency trials for individuals with DS (Nash & Snowling, 2008; Stravroussi et al., 2016).

Limited verbal ability was the leading reason why participants were unable to complete the verbal fluency tasks. It is well-documented that expressive verbal abilities are generally impaired in individuals with DS, highlighting a limitation of this assessment for use across a broad range of participants within this population (Abbeduto et al., 2005; Fidler, 2005; Nash & Snowling, 2008; Stravroussi et al., 2016). However, there is still a need to evaluate verbal measures to provide a wider scope of validated outcome measures for future clinical trials involving children and adolescents with DS.

*Performance on the Verbal Fluency Tasks*

Floor effects were problematic for all trials attempted by our full sample, but this was especially true for the phonemic verbal fluency trials. At least in part, the prevalence of floor effects on the phonemic fluency trials are likely the result of insufficient word retrieval strategies

and other deficits of executive function, as noted in previous research evaluating verbal fluency tasks in children with DS (Nash & Snowling, 2008; Pennington et al., 2003). The high prevalence of motor speech disorders in individuals with DS may also contribute to floor effects on both fluency trials (Wilson et al., 2019). Floor effects are an area of concern for potential outcome measures in DS as they prevent the evaluation of within-group differences and longitudinal change over time. Given that the semantic verbal fluency task nearly met study criteria, our findings suggest that the semantic trials might be a more appropriate measure than the phonemic trials for children and adolescents with DS older than 10 years of age. In addition, use of this measure in future clinical trials is recommended to be limited to including children with verbal abilities.

In addition to floor effects, we also evaluated the effects of the adaptations implemented in scoring task performance. We first examined the number of participants who repeated the examples (i.e., Cat/Dog) provided by the examiners during the task instructions. Approximately 13.7% of participants repeated the examples as their only correct response(s) on the phonemic trials (Bus/Box, Tree/Top), while no participants repeated the examples as their only correct responses on the semantic trials (Cat/Dog, Pizza/Milk). The prevalence of novel responses suggests that participants with DS understand the verbal fluency task instructions and are capable of retrieving words in a given category. While this scoring adaptation does reduce floor effects and enhance the feasibility rate for the phonemic task, even with adapted scoring it still did not meet *a priori* criterion for acceptable feasibility. We then evaluated the inclusion of proper nouns as correct responses on the phonemic /B/ and /T/ verbal fluency trials. One-fifth of participants used proper nouns while generating responses on the /B/ and /T/ trials, yet the impact on performance was only marginally improved. This finding suggests that the inclusion of proper

nouns as correct responses may provide a support for the completion of the phonemic trials, modestly but incompletely improving feasibility.

Consistent with previous research evaluating verbal fluency in DS, sex-related effects were not observed on the semantic and phonemic verbal fluency task (Del Hoyo, 2015). Children and adolescents with DS with higher KBIT-2 raw scores obtained significantly higher scores on most verbal fluency trials as documented in previous normative findings (Liogier d'Ardhuy et al., 2015). Correlations between measures of verbal/nonverbal cognition and perseverations were inconsistent, warranting replication to further understand these associations.

*Test-retest Reliability and Practice Effects*

Test-retest reliability is critically important for outcome measures used in therapeutic trials in DS, where change is assessed from pre- to post-intervention. If a measure is unreliable, it may obscure any real effects that occur in such repeated-measure studies (Edgin et al., 2017; Esbensen et al., 2017; Hessl et al., 2016). The number of correct responses produced on the semantic and the phonemic /T/ fluency trials demonstrated moderate-to-good test-retest reliability, and the phonemic /B/ fluency trial demonstrated poor test-retest reliability. Lower ICC scores were observed on the phonemic fluency trials. These lower ICC scores were expected due to the more challenging nature of the phonemic task for children and adolescents with DS, and the high number of children who were unable to complete the test or received floor scores. Moderate test-retest reliability was observed for the number of perseverations on the semantic animal trials and the number of intrusions on the phonemic /T/ trial (ICC = 0.62 – .67, though poor test-retest reliability was observed for the number of intrusions and perseverations produced during the remaining trials (ICC < .50). Although intrusion and perseveration errors were prevalent, these low-to-moderate ICC scores raise concerns about using intrusions and

perseverations as measures of meaningful change across time. Significant practice effects were not observed on any verbal fluency trial across Time 1 and Time 2. On the surface, these findings might suggest that the verbal fluency tasks would be sensitive to intervention-specific change in future clinical trials involving children and adolescents with DS. However, floor effects could have overwhelmed any potential benefit from practice and could be similarly difficult to overcome by the benefits of an intervention.

*Utility in Children with DS 10 Years and Older*

Though the feasibility of the verbal fluency tasks did not meet *a priori* criteria, sensitivity and specificity probability calculations were used to determine who can complete the measure based on benchmarks of chronological age and IQ. Specificity increases as participant age and IQ increase. High specificity inherently decreases sample size due to a higher exclusion rate, but it also increases the probability that participants will be able to complete the task. By contrast, sensitivity increases as age and IQ restrictions decrease. High sensitivity increases sample size due to a lower exclusion rate, and it consequently grants observation of greater heterogeneity of performance. The semantic verbal fluency trials demonstrate adequate sensitivity and specificity, as well as strong psychometric properties when completed by children with DS 10 years of age and older, although the phonemic verbal fluency trials did not meet *a priori* criteria when completed by the age-restricted sample. The correct responses for the semantic verbal fluency test-retest reliability were also below study criteria, although it was in the upper range of the moderate category (ICC = .72 – .73). Results suggest that the semantic verbal fluency trials are more appropriate for use in future clinical trials enrolling youth with DS 10 years of age and older. It is likely that the phonemic verbal fluency task would only be appropriate for children above 10 years old with higher levels of cognitive ability. Depending on the goals of future

research in DS, the sensitivity and specificity calculations used in the present study can inform inclusion criteria for future clinical trials using verbal fluency outcome measures.

*Limitations and Future Directions*

Despite providing crucial information regarding the usability of verbal fluency tasks for participants with DS, there are study limitations that warrant discussion. The current study was comprised of a racially and ethnically limited sample that should be diversified to better describe performance and to ensure the appropriateness of the task for future use in clinical research with children and adolescents with DS. Similarly, the study had a limited sample of children with DS younger than 10 years of age ($n$ = 22); future work is needed to confirm our findings in a larger sample. Additionally, the test-retest window in the current study was short and it would be beneficial to evaluate the psychometric properties of the assessment, as well as its ability to adequately measure within-individual and within-sample sensitivity to change, across longer time periods (i.e., 12 weeks to 1 year) as would be necessary for future clinical trials in DS. Lastly, ecological validity of verbal fluency tasks was not addressed, and future research is needed to compare performance on verbal fluency tasks with real-world functioning.

To further examine the appropriateness of verbal fluency tasks for future clinical trials involving children and adolescents with DS, future research should describe the convergent validity of verbal fluency tasks with other psychometrically sound outcome measures used to measure executive functioning and language in children and adolescents with DS. Importantly, investigators should focus on measures with known ecological validity to determine the relationship between verbal fluency task performance and everyday functioning. Recent research has identified sound measures of executive function and language for children with DS, such as the Behavior Rating Inventory of Executive Function, 2nd Edition and expressive language

sampling, though there is still a need for the identification of measures with a known connection to daily performance to increase confidence in the ability of verbal fluency task to capture real-world changes in future clinical trials (Chaytor & Schmitter-Edgecombe, 2003; Esbensen et al., 2019; Thurman et al., 2021). Further, future work should evaluate the use of visual supports for the completion of the verbal fluency task (i.e., showing participants the letter /B/ or /T/ prior to administering the phonemic verbal fluency trials), to determine whether added visual supports increase the feasibility of the task in children and adolescents with DS. Lastly, clustering and switching coding processes provide critical information about how responses are generated during the task's administration that are not collected by measuring the number of correct and incorrect responses alone and should be explored in future studies (Nash & Snowling, 2008; Troyer, 1997; Troyer, 1998; Troyer, 2000).

*Conclusion*

When completed by the entire sample, the verbal fluency tasks failed to display sound psychometric properties, though the semantic fluency trials had stronger psychometric indices than the phonemic trials. The evaluation of an age-restricted sample supported the use of the semantic verbal fluency task for children with DS 10 years of age and older, however, test-retest reliability was problematic. The phonemic task may only be appropriate for participants with DS who have advanced cognitive skills regardless of age. Though error measurements produced on the task have previously been used to further evaluate cognitive processes during performance on the task, the current study revealed a lack of stability for these measures to provide meaningful information across time. Given the similarity of administration and scoring procedures, we believe that the results from the present study can be generalized to most verbal fluency outcome measures completed by children and adolescents with DS, though further research should be

conducted to validate this claim. This study supports the continued selection of outcome measures for future therapeutic trials in DS and other populations with intellectual disabilities.

**REFERENCES**

Abbeduto, L., Chapman, R. S., Fletcher, P., & Miller, J. (2005). Language development in Down syndrome and fragile X syndrome. *Trends in Language Acquisition Research, 4*, 53-72.

Abbeduto, L., Warren, S. F., & Conners, F. A. (2007). Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(3), 247-261.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91-93.

Azuma, T. (2004). Working memory and perseveration in verbal fluency. *Neuropsychology, 18*(1), 69-77. doi:10.1037/0894-4105.18.1.69

Ball, S. L., Holland, A. J., Treppner, P., Watson, P. C., & Huppert, F. A. (2008). Executive dysfunction and its association with personality and behaviour changes in the development of Alzheimer's disease in adults with Down syndrome and mild to moderate learning disabilities. *British Journal of Clinical Psychology, 47*(1), 1-29.

Becker, N., & Salles, J. F. d. (2016). Methodological criteria for scoring clustering and switching in verbal fluency tasks. *Psico-usF, 21*(3), 445-457.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327-336.

Charchat-Fichman, H., Oliveira, R. M., & Silva, A. M. d. (2011). Performance of Brazilian children on phonemic and semantic verbal fluency tasks. *Dementia & Neuropsychologia, 5*(2), 78-84.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197.

Costafreda, S. G., Fu, C. H., Lee, L., Everitt, B., Brammer, M. J., & David, A. S. (2006). A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. *Human Brain Mapping, 27*(10), 799-810.

Daunhauer, L. A., Fidler, D. J., & Will, E. (2014). School function in students with Down syndrome. *American Journal of Occupational Therapy, 68*(2), 167-176.

de la Torre, R., de Sola, S., Hernandez, G., Farré, M., Pujol, J., Rodriguez, J., . . . Principe, A. (2016). Safety and efficacy of cognitive training plus epigallocatechin-3-gallate in young adults with Down's syndrome (TESDAD): a double-blind, randomised, placebo-controlled, phase 2 trial. *The Lancet Neurology, 15*(8), 801-810.

Del Hoyo Soriano, L., Xicota Vila, L., Sánchez Benavides, G., Cuenca Royo, A. M., Sola Llopis, S. d., Langohr, K., . . . Fornell, T. (2015). Semantic Verbal Fluency Pattern, Dementia Rating Scores and Adaptive Behavior Correlate With Plasma Aβ42 Concentrations in Down Syndrome Young Adults. *Frontiers in Behavioral Neuroscience*, 9, 301.

Edgin, J. O., Anand, P., Rosser, T., Pierpont, E. I., Figueroa, C., Hamilton, D., . . . Toole, L. (2017). The Arizona Cognitive Test Battery for Down syndrome: Test-retest reliability and practice effects. *American Journal on Intellectual and Developmental Disabilities, 122*(3), 215-234.

Edgin, J. O., Mason, G. M., Allman, M. J., Capone, G. T., DeLeon, I., Maslen, C., . . . Nadel, L. (2010). Development and validation of the Arizona Cognitive Test Battery for Down syndrome. *Journal of Neurodevelopmental Disorders, 2*(3), 149-164.

Esbensen, A. J., Hoffman, E. K., Shaffer, R., Chen, E., Patel, L., & Jacola, L. (2019). Reliability of informant-report measures of executive functioning in children with down syndrome. American journal on intellectual and developmental disabilities, 124(3), 220-233.

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S. L., Edgin, J., d'Ardhuy, X. L., . . . Abbeduto, L. (2017). Outcome measures for clinical trials in Down syndrome. *American Journal on Intellectual and Developmental Disabilities, 122*(3), 247-281.

Fabbretti, D., Pizzuto, E., Vicari, S., & Volterra, V. (1997). A story description task in children with Down's syndrome: Lexical and morphosyntactic abilities. *Journal of Intellectual Disability Research, 41*(2), 165-179.

Fidler, D. J. (2005). The emerging Down syndrome behavioral phenotype in early childhood: Implications for practice. *Infants & Young Children, 18*(2), 86-103.

Fidler, D. J., Daunhauer, L. A., Schworer, E., & Patel, L. (2020). Executive function in Down syndrome: Links to adaptation and treatment implications. In *The Oxford Handbook of Down Syndrome and Development*.

Hamburg, S., Rosch, R., Startin, C. M., Friston, K. J., & Strydom, A. (2019). Dynamic Causal Modeling of the Relationship between Cognition and Theta–alpha Oscillations in Adults with Down Syndrome. *Cerebral Cortex, 29*(5), 2279-2290.

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test- retest reliability. *British Journal of Clinical Psychology, 39*(2), 181-191.

Hart, S. J., Visootsak, J., Tamburri, P., Phuong, P., Baumer, N., Hernandez, M. C., . . . Kishnani, P. S. (2017). Pharmacological interventions to improve cognition and adaptive

functioning in Down syndrome: Strides to date. *American Journal of Medical Genetics Part A, 173*(11), 3029-3041.

Heller, J. H., Spiridigliozzi, G. A., Crissman, B. G., Sullivan- Saarela, J. A., Li, J. S., & Kishnani, P. S. (2006). *Clinical trials in children with Down syndrome: issues from a cognitive research perspective.* Paper presented at the American Journal of Medical Genetics Part C: Seminars in Medical Genetics.

Hessl, D., Sansone, S. M., Berry-Kravis, E., Riley, K., Widaman, K. F., Abbeduto, L., . . . Rhodes, K. C. (2016). The NIH Toolbox Cognitive Battery for intellectual disabilities: three preliminary studies and future directions. *Journal of Neurodevelopmental Disorders, 8*(1), 35.

Hewitt, L. E., Hinkle, A. S., & Miccio, A. W. (2005). Intervention to improve expressive grammar for adults with Down syndrome. *Communication Disorders Quarterly, 26*(3), 144-155.

Hurks, P., Vles, J., Hendriksen, J., Kalff, A., Feron, F., Kroes, M., . . . Jolles, J. (2006). Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *Journal of Clinical and Experimental Neuropsychology, 28*(5), 684-695.

Kaufman, A. S., & Kaufman, N. L. (2004). *Manual: Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Keeling, L. A., Spiridigliozzi, G. A., Hart, S. J., Baker, J. A., Jones, H. N., & Kishnani, P. S. (2017). Challenges in measuring the effects of pharmacological interventions on cognitive and adaptive functioning in individuals with Down syndrome: A systematic review. *American Journal of Medical Genetics Part A, 173*(11), 3058-3066.

Kishnani, P. S., Heller, J. H., Spiridigliozzi, G. A., Lott, I., Escobar, L., Richardson, S., . . . McRae, T. (2010). Donepezil for treatment of cognitive dysfunction in children with Down syndrome aged 10–17. *American Journal of Medical Genetics Part A, 152*(12), 3028-3035.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY II: Clinical and interpretive manual*: Harcourt Assessment, PsychCorp.

Kumin, L., Councill, C., & Goodman, M. (1994). A longitudinal study of the emergence of phonemes in children with Down syndrome. *Journal of Communication Disorders, 27*(4), 293-303.

Laws, G. (2004). Contributions of phonological memory, language comprehension and hearing to the expressive language of adolescents and young adults with Down syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1085-1095.

Liogier d'Ardhuy, X., Edgin, J. O., Bouis, C., de Sola, S., Goeldner, C., Kishnani, P., . . . Squassante, L. (2015). Assessment of cognitive scales to examine memory, executive function and language in individuals with Down syndrome: Implications of a 6-month observational study. *Frontiers in Behavioral Neuroscience, 9*, 300.

Marcell, M. M. (1995). Relationships between hearing and auditory cognition in Down's syndrome youth. *Downs Syndrome Research and Practice*, *3*, 75-91.

Memisevic, H., Biscevic, I., & Pasalic, A. (2017). Developmental trends in semantic fluency in preschool children. *Cogent Psychology, 4*(1), 1403064.

Næss, K.-A. B., Lyster, S.-A. H., Hulme, C., & Melby-Lervåg, M. (2011). Language and verbal short-term memory skills in children with Down syndrome: A meta-analytic review. *Research in Developmental Disabilities, 32*(6), 2225-2234.

Nash, H. M., & Snowling, M. J. (2008). Semantic and phonological fluency in children with Down syndrome: Atypical organization of language or less efficient retrieval strategies? *Cognitive Neuropsychology, 25*(5), 690-703.

Parker, S. E., Mai, C. T., Canfield, M. A., Rickard, R., Wang, Y., Meyer, R. E., . . . Kirby, R. S. (2010). Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006. *Birth Defects Research Part A: Clinical and Molecular Teratology, 88*(12), 1008-1016.

Pennington, B. F., Moon, J., Edgin, J., Stedron, J., & Nadel, L. (2003). The neuropsychology of Down syndrome: evidence for hippocampal dysfunction. *Child Development, 74*(1), 75-93.

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. Journal of clinical and experimental neuropsychology, 20(3), 310-319.

Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language, 71*(2), 267-284.

Sabat, C., Arango, P., Tassé, M. J., & Tenorio, M. (2020). Different abilities needed at home and school: The relation between executive function and adaptive behaviour in adolescents with Down syndrome. *Scientific Reports, 10*(1), 1-10.

Schworer, E.K., Esbensen, A.J., Fidler, D.J., Beebe, D.W., Carle, A., & Wiley, S. (2021). Evaluating working memory outcome measures for children with Down syndrome. Journal of Intellectual Disability Research. doi: 10.1111/jir.12833

Snow, W., Tierney, M., Zorzitto, M., Fisher, R., & Reid, D. (1988). *One-year test-retest reliability of selected neuropsychological tests in older adults.* Paper presented at the Journal of Clinical and Experimental Neuropsychology.

Snowling, M. J., Hulme, C., & Mercer, R. C. (2002). A deficit in rime awareness in children with Down syndrome. *Reading and Writing*, *15*(5-6), 471-495.

Stavroussi, P., Andreou, G., & Karagiannopoulou, D. (2016). Verbal fluency and verbal short-term memory in adults with Down syndrome and unspecified intellectual disability. *International Journal of Disability, Development and Education, 63*(1), 122-139.

Thurman, A. J., Edgin, J. O., Sherman, S. L., Sterling, A., McDuffie, A., Berry-Kravis, E., Hamilton, D., & Abbeduto, L. (2021). Spoken language outcome measures for treatment studies in Down syndrome: feasibility, practice effects, test-retest reliability, and construct validity of variables generated from expressive language sampling. *Journal of neurodevelopmental disorders, 13*(1), 1-17.

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 370-378.

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.

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society, 4*(2), 137-143.

Will, E.A., Schworer, E.K., & Esbensen, A.J. (2021) The role of distinct executive functions on adaptive behavior in children and adolescents with Down syndrome. Child Neuropsychology. doi:10.1080/09297049.2021.1917531

Wilson, E. M., Abbeduto, L., Camarata, S. M., & Shriberg, L. D. (2019). Speech and motor speech disorders and intelligibility in adolescents with Down syndrome. Clinical linguistics & phonetics, 33(8), 790-814.

Wright, C. A., Kaiser, A. P., Reikowsky, D. I., & Roberts, M. Y. (2013). Effects of a naturalistic sign intervention on expressive language of toddlers with Down syndrome. *Journal of Speech, Language, and Hearing Research*.

Table 1
*Feasibility of the verbal fluency tasks at Time 1 and Time 2, n = 85*

| | Feasibility n (%) | Intrusions n (%) | Perseverations n (%) |
|---|---|---|---|
| **Semantic Trial: Animal** | 64 (75.3) | 29 (34.1) | 47 (55.3) |
| **Semantic Trial: Food** | 63 (74.1) | 17 (26.9) | 32 (50.8) |
| **Phonemic Trial: /B/** | 51 (60.0) | 31 (60.8) | 26 (51.0) |
| **Phonemic Trial: /T/** | 50 (58.8) | 30 (60.0) | 25 (50.0) |

*Note.* Intrusions n = number of participants with intrusions at Time 1, Perseverations n = number of participants with perseverations at Time 1

*Table 2. Performance and associations with sex, age, and IQ at Time 1*

| | Mean (SD) | Median (Min, Max) | Skew | Kurtosis | Sex Diff (*t/U*) | Correlations | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Age | KBIT-2 V | KBIT-2 NV |
| ***Semantic Trial: Animal***, n = 64 | | | | | | | | |
| Total correct $^{r}$ | 8.94 (4.42) | 9.00 (1, 20) | .38 | -.26 | 1.01 | .23 | .50** | .52** |
| Number of Intrusions $^{\rho, U}$ | .80 (1.14) | .00 (0,6) | 2.00 | 5.64 | 446.50 | .00 | -.02 | -.11 |
| Number of Perseverations $^{\rho, U}$ | 2.16 (2.21) | 1.00 (0,8) | 1.07 | .41 | 478.00 | -.32** | -.14 | -.07 |
| ***Semantic Trial: Food***, n = 63 | | | | | | | | |
| Total correct $^{r}$ | 10.14 (4.18) | 9.00 (3, 19) | .34 | -.59 | -.10 | .17 | .37** | .39** |
| Number of Intrusions $^{\rho, U}$ | .41 (.85) | .00 (0,5) | 3.08 | 12.75 | 414.50 | .01 | -.07 | -.01 |
| Number of Perseverations $^{\rho, U}$ | 1.21 (1.60) | 1.00 (0,7) | 1.59 | 2.62 | 458.50 | -.24 | -.08 | -.15 |
| ***Phonemic Trial: /B/***, n = 51 | | | | | | | | |
| Total correct $^{r}$ | 5.31 (2.21) | 5.00 (2, 12) | .70 | .62 | -.14 | -.10 | .16 | .28* |
| Number of Intrusions $^{\rho, U}$ | 2.20 (2.71) | 1.00 (0,9) | 1.29 | .60 | 312.50 | .09 | .04 | -.12 |
| Number of Perseverations $^{\rho, U}$ | 1.06 (1.53) | 1.00 (0,6) | 1.72 | 2.25 | 255.00 | -.20 | .32* | .25 |
| ***Phonemic Trial: /T/***, n = 50 | | | | | | | | |
| Total correct $^{r}$ | 4.90 (2.41) | 5.00 (0, 12) | .75 | .35 | .13 | .23 | .12 | .41* |
| Number of Intrusions $^{\rho, U}$ | 2.14 (2.72) | 1.00 (0, 11) | 1.50 | .35 | 77.00 | -.27 | -.17 | -.03 |
| Number of Perseverations $^{\rho, U}$ | 1.02 (1.30) | .50 (0, 5) | 1.18 | .58 | 72.00 | -.30 | .29 | .18 |

*Note.* Sample includes participants who completed the task at Time 1 and Time 2.
*$p<.05$, **$p<.01$; $^{r}$: bivariate Pearson correlations, $^{\rho}$: Spearman ρ correlations; *U:* Mann Whitney U T-Test;
*KBIT-2* = Kaufman Brief Intelligence Test, second edition; KBIT-2 V = KBIT-2 Verbal raw score, KBIT-2 NV = KBIT-2 Nonverbal raw score.

Table 3.
*Examination of test-retest reliability and practice effects*

| | Time 1 Mean (SD) | Time 2 Mean (SD) | Mean Difference | *t* | Cohen's *d* | ICC |
|---|---|---|---|---|---|---|
| ***Semantic Trial: Animal,*** n = 64 | | | | | | |
| **Total correct** | 8.94 (4.42) | 8.50 (4.19) | .44 | 1.10 | .10 | .72 |
| **Number of Intrusions** | .80 (1.14) | .89 (1.49) | -.09 | -.49 | .07 | .34 |
| **Number of Perseverations** | 2.16 (2.21) | 2.08 (1.99) | .08 | .28 | .04 | .62 |
| ***Semantic Trial: Food,*** n = 63 | | | | | | |
| **Total correct** | 10.14 (4.18) | 9.46 (4.46) | .68 | 1.80 | .16 | .75 |
| **Number of Intrusions** | .41 (.85) | .48 (.64) | -.07 | -.61 | .09 | .41 |
| **Number of Perseverations** | 1.21 (1.60) | 1.63 (1.64) | .42 | -2.06 | .26 | .47 |
| ***Phonemic Trial: /B/,*** n = 51 | | | | | | |
| **Total correct** | 5.38 (2.18) | 5.20 (2.41) | .18 | .51 | .08 | .42 |
| **Number of Intrusions** | 2.18 (2.74) | 1.96 (2.47) | .22 | .59 | .08 | .49 |
| **Number of Perseverations** | 1.00 (1.49) | 1.56 (2.60) | -.56 | -1.57 | .27 | .29 |
| ***Phonemic Trial: /T/,*** n = 50 | | | | | | |
| **Total correct** | 4.90 (2.41) | 4.90 (2.65) | .00 | .00 | .00 | .60 |
| **Number of Intrusions** | 2.14 (2.72) | 1.96 (2.76) | .18 | .57 | .07 | .67 |
| **Number of Perseverations** | 1.02 (1.30) | 1.06 (2.24) | -.04 | -.12 | .02 | .37 |

*Note.* Sample includes participants who completed the task at Time 1 and Time 2.

Table 4.
*Sensitivity and specificity benchmarks for semantic and phonemic verbal fluency trials*

| | **Semantic Trials** | | **Phonemic Trials** | |
|---|---|---|---|---|
| | *Sensitivity* | *Specificity* | *Sensitivity* | *Specificity* |
| **No FSIQ restriction (age > 10, *n* = 63)** | 92% | 51% | 92% | 51% |
| **Age and FSIQ above 40 (*n* =29)** | 43% | 91% | 50% | 89% |
| **Age and FSIQ above 45 (*n* =18)** | 25% | 91% | 30% | 91% |
| **Age and FSIQ above 50 (*n* = 12)** | 17% | 95% | 20% | 94% |

Table 5.
*Feasibility of the verbal fluency tasks for participants 10 years and older at Time 1 and Time 2, n = 63*

| | Feasibility n (%) | Intrusions n (%) | Perseverations n (%) |
|---|---|---|---|
| **Semantic Trial: Animal** | 53 (84.1) | 25 (47.2) | 38 (71.7) |
| **Semantic Trial: Food** | 53 (84.1) | 17 (32.1) | 24 (45.3) |
| **Phonemic Trial: /B/** | 46 (73.0) | 28 (60.9) | 24 (52.2) |
| **Phonemic Trial: /T/** | 46 (73.0) | 27 (658.7) | 23 (50.0) |

*Note.* Intrusions n = number of participants with intrusions at Time 1, Perseverations n = number of participants with perseverations at Time 1

Table 6.

*Examination of test-retest reliability and practice effects for participants 10 years and older*

| | Time 1 Mean (SD) | Time 2 Mean (SD) | Mean Difference | *t* | Cohen's *d* | ICC |
|---|---|---|---|---|---|---|
| ***Semantic Trial: Animal,*** n = 53 | | | | | | |
| **Total correct** | 9.62 (4.31) | 9.00 (4.18) | .62 | 1.43 | .15 | .72 |
| **Number of Intrusions** | .87 (1.21) | .92 (1.59) | -.06 | -.26 | .04 | .36 |
| **Number of Perseverations** | 1.94 (2.03) | 2.08 (2.07) | -.13 | -.43 | .07 | .40 |
| ***Semantic Trial: Food,*** n = 53 | | | | | | |
| **Total correct** | 10.66 (4.10) | 10.28 (4.32) | .38 | .89 | .09 | .73 |
| **Number of Intrusions** | .49 (.91) | .47 (.67) | .02 | .16 | .03 | .45 |
| **Number of Perseverations** | .98 (1.4) | 1.53 (1.59) | -.55 | -2.4 | .37 | .52 |
| ***Phonemic Trial: /B/,*** n = 46 | | | | | | |
| **Total correct** | 5.41 (2.25) | 5.26 (2.44) | .15 | .41 | .06 | .42 |
| **Number of Intrusions** | 2.22 (2.80) | 2.02 (2.55) | .20 | .49 | .07 | .50 |
| **Number of Perseverations** | 1.02 (1.53) | 1.67 (2.68) | -.65 | -1.71 | .30 | .29 |
| ***Phonemic Trial: /T/,*** n = 46 | | | | | | |
| **Total correct** | 5.09 (2.39) | 4.98 (2.72) | .11 | .32 | .04 | .59 |
| **Number of Intrusions** | 2.20 (2.81) | 2.07 (2.85) | .13 | .35 | .05 | .69 |
| **Number of Perseverations** | 1.04 (1.33) | .96 (2.25) | .09 | .40 | .04 | .28 |

*Note.* Sample includes participants who completed the task at Time 1 and Time 2.

Table 7.
*Summary of reported subtest quality based on a priori criteria*

| | Minimal floor effects | Feasibility | Test-retest | Negligible practice effects |
|---|---|---|---|---|
| ***Total sample, n = 85*** | | | | |
| **Semantic Trial: Animal** | – | – | – | + |
| **Semantic Trial: Food** | – | – | + | + |
| **Phonemic Trial: /B/** | – | – | – | + |
| **Phonemic Trial: /T/** | – | – | – | + |
| ***>10-year-old sample, n = 63*** | | | | |
| **Semantic Trial: Animal** | + | + | – | + |
| **Semantic Trial: Food** | + | + | – | + |
| **Phonemic Trial: /B/** | – | – | – | + |
| **Phonemic Trial: /T/** | – | – | – | + |

+ indicates study criterion met: < 20% floor effects, ≥ 80% feasibility, ≥ .75 test-restest ICC, small and non-significant practice effects; – indicates study criterion not met: ≥ 20% floor effects, < 80% feasibility, < .75 test-retest ICC, medium/large and significant practice effects.

Figure 1.
Chronological age and KBIT-2 FSIQ of completers and non-completers for semantic verbal fluency trials

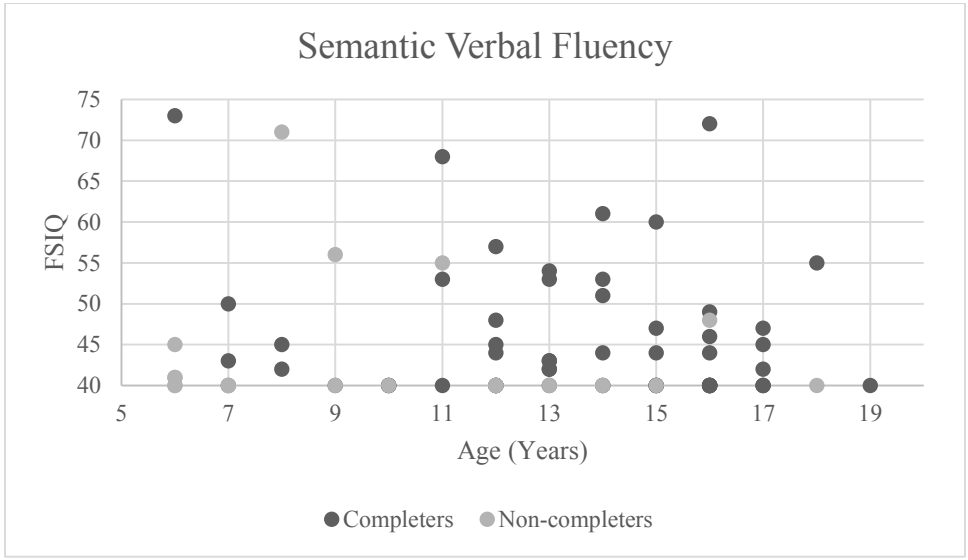

Figure 2.
Chronological age and KBIT-2 FSIQ of completers and non-completers for phonemic verbal fluency trials