Title

Psychometric evaluation of a working memory assessment measure in young children with Down syndrome

Authors
Pinks, Miranda E
Van Deusen, Kaylyn
Prince, Mark A
et al.

Publication Date
2023-08-01

DOI
10.1016/j.ridd.2023.104564

Peer reviewed

HHS Public Access
Author manuscript
Res Dev Disabil. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:
Res Dev Disabil. 2023 August ; 139: 104564. doi:10.1016/j.ridd.2023.104564.

Psychometric evaluation of a working memory assessment measure in young children with Down syndrome

Miranda E. Pinks[a], **Kaylyn Van Deusen**[a], **Mark A. Prince**[b], **Anna J. Esbensen**[c,d], **Angela John Thurman**[e], **Lina R. Patel**[f], **Leonard Abbeduto**[e], **Madison M. Walsh**[a], **Lisa A. Daunhauer**[a], **Robyn Tempero Feigles**[e], **Vivian Nguyen**[e], **Deborah J. Fidler**[a]

[a]Human Development and Family Studies, Colorado State University, Fort Collins, CO, USA

[b]Department of Psychology, Colorado State University, Fort Collins, CO, USA

[c]Division of Developmental and Behavioral Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[d]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

[e]Department of Psychiatry, MIND Institute, University of California Davis Health, Sacramento, CA, USA

[f]Department of Psychiatry, University of Colorado-Anschutz Medical Campus, Aurora, CO, USA

Abstract

Background: Working memory involves the temporary storage and manipulation of information and is frequently an area of challenge for individuals with Down syndrome (DS). Despite the potential benefits of intervention, laboratory assessments of working memory that could capture intervention effects have not undergone rigorous evaluation for use with young children with DS. It is critical to evaluate assessments of working memory in young children with DS to ensure the reliable and accurate measurement of performance.

Aim: This study evaluated an adapted laboratory measure of working memory for young children with DS 2–8 years old.

Method: A self-ordered pointing task, the *Garage Game*, was administered to 78 children with DS (mean = 5.17 years; SD = 1.49). Adaptations were made to the task to minimize potential DS phenotyperelated language and motor confounds.

Author Contributions:
Miranda E. Pinks: Formal analysis, Writing - Original Draft, Visualization. **Kaylyn Van Deusen:** Investigation, Data curation, Writing - Review & Editing, Project administration. **Mark A. Prince:** Methodology, Formal Analysis, Visualization. **Anna J. Esbensen:** Supervision, Writing - Review & Editing. **Angela John Thurman:** Supervision, Writing - Review & Editing. **Lina R. Patel:** Investigation, Writing - Review & Editing. **Leonard Abbeduto:** Supervision, Writing - Review & Editing. **Madison M. Walsh:** Investigation, Data Curation. **Lisa A. Daunhauer:** Conceptualization. **Robyn Tempero Feigles:** Investigation. **Vivian Nguyen:** Investigation. **Deborah J. Fidler:** Conceptualization, Methodology, Resources, Writing - Original Draft, Supervision, Funding Acquisition.

Results: Results indicate that the measure is feasible, scalable, and developmentally sensitive, with minimal floor and practice effects for this population within this chronological and developmental age range.

Conclusion: These findings demonstrate that the *Garage Game* is promising for use in studies of early working memory and treatment trials that aim to support the development of this critical dimension of executive functioning for children with DS.

Keywords

Down syndrome; assessment; working memory

1. Introduction

Down syndrome (DS) is a neurogenetic condition that occurs in approximately 1 in 700 births (Mai et al., 2019). Individuals with DS are predisposed to develop cognitive and behavioral strengths and challenges that reflect a distinct phenotypic profile (Fidler, 2005; Grieco et al., 2015). One area in which people with DS display both relative strengths and weaknesses is the self-regulatory goal-directed set of skills called ‗executive function' (EF; Daunhauer & Fidler, 2013). Individuals with DS often demonstrate challenges with EF compared with peers of the same chronological age (CA; Danielsson et al., 2010; Rowe et al., 2006) and mental age (MA; Daunhauer et al., 2014; Lanfranchi et al., 2010). This paper aims to evaluate a performance-based measure of a particular EF component, working memory, for applications in future research and clinical studies involving young children with DS.

EF is typically conceptualized as a unitary construct with dissociable components (Miyake et al., 2000). Most models of EF include the components of working memory, inhibition, and cognitive flexibility (Best & Miller, 2010; Blair et al., 2005; Garon et al., 2008; Miyake et al., 2000). Working memory, the ability to temporarily store and manipulate information, plays an important role in everyday tasks such as reading, learning, and problemsolving (Baddeley, 1983). Although challenges in verbal short-term memory are well-established in individuals with DS (Jarrold, 2001), relatively less emphasis has been placed on more domaingeneral working memory skills. However, across almost all studies that examine working memory in DS, it has emerged as an area of relative challenge, with more pronounced delays than expected based on overall developmental status (see Daunhauer et al., 2014, 2017; Tungate & Conners, 2021; Van Deusen et al., 2022). In a recent exploration of EF profiles among 69 children with DS aged 3–10, two profiles emerged: one with elevated challenges in working memory only, and one with elevated challenges in working memory, inhibition, and planning (Van Deusen et al., 2022). These findings are notable in that both profiles involved pronounced challenges in working memory regardless of the presence or absence of other EF challenges (Van Deusen et al., 2022). A separate analysis of EF in individuals 2–35 years of age with DS also reported working memory as a principal area of difficulty beginning in the preschool years and continuing through adolescence (Loveall et al., 2017).

Intervention research that supports the development of working memory has become a priority in DS due to its relevance for academic achievement and employment (Tomaszewski et al., 2018; Will et al., 2017). Behavioral interventions such as computerized training have shown promise in improving working memory, especially in the visuospatial domain (Bennett et al., 2013; Lanfranchi et al., 2017). Despite the potential benefits of these interventions, the laboratory assessments used to measure working memory in these studies have not undergone rigorous psychometric evaluation in populations with DS. This highlights the need for laboratory measures of working memory skills in individuals with DS because a lack of appropriate measures in clinical trials could lead to erroneous conclusions about the effectiveness of interventions.

1.1 Measuring EF in young children with DS

In research on DS, additional considerations are warranted for the assessment and measurement of cognitive skills. For example, EF tasks that place a high demand on language processing or motor skills (often areas of relative difficulty in children with DS) may not accurately capture the intended construct and may yield results confounded by performances in areas unrelated to EF. In 2015, the National Institutes of Health (NIH) assembled a working group to identify potentially appropriate cognitive measures for individuals with DS (Esbensen et al., 2017). This group expressed the need for psychometrically adequate measures appropriate for individuals with DS that could be used to detect treatment effects in clinical trials (Esbensen et al., 2017). Since that time, researchers have worked to establish the psychometric properties of measures for individuals with DS, including assessments of, among other things, cognitive flexibility (Schworer et al., 2023), verbal fluency (Smeyne et al., 2022), social cognition and behavior (Schworer et al., 2021), and neurocognitive function (Edgin et al., 2017). One recent study (Schworer et al., 2022) demonstrated that working memory subtests from various standardized clinical assessments, including Recall of Sequential Order, Recall of Digits Forward and Recall of Digits Backward, and Recognition of Pictures from the Differential Ability Scales, Second Edition (DAS-2; Elliott, 2007), the verbal and non-verbal working memory subtests of the Stanford Binet Intelligence Scales, Fifth Edition (SB-5; Roid, 2003b), and the Picture Span task from the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-5; Wechsler, 2014), met most psychometric criteria for individuals with DS in the chronological age range of 6 to 17 years (Schworer et al., 2022). However, none of the measures in this comprehensive evaluation were tested in young children with DS and many of these measures are known to demonstrate floor effects in children with young mental ages (e.g., Sansone et al., 2014). Consequently, the ability to accurately evaluate and detect changes in working memory performances in children with DS under 6 years of age remains an important gap in our current measurement resources.

1.2 Working memory: An important aspect of EF

Laboratory assessments of working memory in DS have often involved reverse span approaches, in which participants are tasked with demonstrating how much visuospatial or verbal information they can temporarily retain and manipulate (Tungate & Conners, 2021). Commonly used assessments of working memory include backward digit span tasks (e.g., Elliott, 2007; Wechsler, 2014) and the backward Corsi task (Isaacs & Vargha-Khadem, 1989), both of which require the participant to remember a span of items (i.e., numbers, spatial sequence) and recall them in reverse order. Another important aspect of working memory, however, is the ability to relate multiple pieces of information that are held in memory together or inhibit distractions while maintaining information in short-term memory (Diamond et al., 1997). Early working memory ability is thought to undergird later academic skills such as reading comprehension and mathematical performance (Swanson & Alloway, 2012) and is, therefore, an important potential target for intervention for children with DS. A measure of working memory for young children with DS may contribute to a broader understanding of the full range of working memory dimensions and will contribute to the literature that emphasizes the verbal component of memory processes in this population. To advance intervention and treatment science in this area, measures are needed that are appropriate for both overall developmental status and consideration of specific developmental features often associated with DS, like language delays and motor planning challenges (Fidler, 2005).

1.3 Importance of appropriate performance-based measures

The preschool years are a time of rapid growth for EF and are considered an especially important period for intervention (Zelazo & Carlson, 2012). Recent calls to advance intervention and treatment for people with DS have emphasized the need for interventions that capitalize on early neuroplasticity (Edgin et al., 2015). Indeed, training programs have shown promise for improving EF skills, including working memory, in older children/adolescents with and without DS (Bennett et al., 2013; Conners et al., 2008; Diamond et al., 2007). However, performancebased measures that are sensitive to early changes in processing efficiency in DS are essential for capturing the treatment effects of EF intervention in early childhood, when treatments may have the greatest downstream effects.

Most studies involving direct assessments of EF in DS include participants who are older in CA (Tungate & Conners, 2021), and few studies have included participants in the early childhood range. For illustration, a recent meta-analysis of 57 studies comparing EF skills in DS to MA-matched groups without DS reported a mean CA of 14.46 and range of 6.62–40.90 years among over 3,000 participants with DS (Tungate & Conners, 2021). Using currently available proxy-report rather than performance measures for the assessment of working memory in young children with DS may provide an incomplete picture of this skill area (Toplak et al., 2013). Therefore, the evaluation of laboratory measures of working memory for young children with DS will contribute to our overall understanding of EF development in children with DS and can facilitate the accurate measurement of incremental changes in raw processing efficiency, including capturing the effects of potential future interventions.

1.4 Current Study

The current study aimed to evaluate a developmentally appropriate, syndrome-informed working memory performance task, the *Garage Game*, for use with children 2.50–8.67 years old with DS. The primary goal of this work was to (1) evaluate the feasibility of this adapted working memory assessment for use in this population both in terms of CA and MA and (2) identify any distributional limits (i.e., ceiling and floor effects) that constrain the utility of the task for children with DS in this CA range. We also aimed to examine the scalability of this measure and explore preliminary test-retest reliability and practice effects. These data are needed to guide recommendations for the use of the *Garage Game* task for future treatment studies that include young children with DS.

2. Method

2.1 Participants

Participants were 78 children aged 2.5–8.67 years (mean CA = 5.17 years; SD = 1.49). with a confirmed DS diagnosis. A subsample of participants from the multi-wave larger study was also included in an examination of test-retest reliability and practice effects two weeks later. These 22 participants had a CA range of 2.67 to 8.67 years (M = 5.88, SD = 1.64). See Table 1 for additional participant demographics. English language understanding was required for inclusion for both child participants and their caregivers.

2.2 Procedures

Participants were drawn from two larger studies of cognition in DS. The Western Institutional Review Board and the IRB at [*withheld for review*] approved the procedures and oversaw each site participating in the project. Caregivers of the participants provided informed consent to participate in the project. Caregivers completed questionnaires, and child participants completed a battery of assessments measuring language, motor skills, developmental status, and EF. All tasks were presented to the children as games. As needed, children were given opportunities for breaks throughout the assessment period. Families received monetary compensation for their participation in the study, with amount approved by the appropriate Institutional Review Board. Study data were collected and managed using REDCap electronic data capture tools hosted at [*withheld for review*] (Harris et al., 2009, 2019).

2.3 Measures

2.3.1 Medical History Questionnaire—Parents were asked to complete a series of questions regarding their child's medical history, including DS type, sensory difficulties (e.g., vision and hearing), biomedical risk factors (e.g., premature birth), and psychological diagnoses (e.g., autism spectrum disorder (ASD) and attentional deficit and/or hyperactivity disorder (ADD/ADHD)). Rates of DS type, premature birth, congenital heart defects, ASD, and ADD/ADHD are reported in Table 1.

2.3.2 Cognitive Status—Two assessments were used to assess cognitive status, the Cognitive domain of the Bayley Scales of Infant and Toddler Development-4 (Bayley-4; Bayley & Aylward, 2019) in children 30–36 months old and the SB-5 Abbreviated Battery IQ (SB-5 ABIQ; Roid, 2003b) in children over 36 months old. The Bayley-4 is a standardized assessment of cognition, communication, and motor skills designed for children 16 days to 42 months old. Standardized scores from the Bayley-4 are normed based on US Census data. The Bayley-4 scales have high internal reliability (.93 to .95) and rest-retest reliability (.81 to .84), even when assessing individuals with different levels of development or clinical diagnoses (Bayley & Aylward, 2019). The SB-5 is a normed and standardized assessment that has become a common data element in DS research and was administered to all children over the age of 3 years (Roid, 2003a). The SB-5 ABIQ has high internal reliability (.91) and test-retest reliability (0.85) reported in the technical manual (Roid, 2003a).

Age equivalent (AE) scores from the SB-5 ABIQ were used for data analysis; however, there were concerns regarding potential floor effects within this sample, as the lowest AE score on the SB-5 is 24 months. To address this concern, when possible, children with a CA of 36 to 59 months in one of the studies were also administered the Cognitive domain scale of the Bayley-4. Due to time limitations and participant fatigue, only 12 of 34 participants aged 36–59 months received both assessments, and one participant in this age range received only the Bayley-4 due to refusal to complete the SB-5 ABIQ. We also note that, though the Bayley-4 is designed for use in children up to 42 months of age, the measure was within the appropriate developmental range for children with DS and age equivalent scores were within range in this sample.

SB-5 ABIQ AE scores were generated for the majority of participants (96%) over 3.0 years old, and 35 (46.7%) had AE scores at the floor. Bayley-4 scores were available for 12 of those at the floor and one participant who was missing a SB-5 ABIQ score. There were 23 children who received a floor AE of 24 months and did not have Bayley-4 scores. To address this censored data issue, MA-related analyses used one-year age bands (i.e., 12–24 months = 1 year; 25–35 months = 2 years; 36–47 months = 3 years; 48–59 months = 4 years). Within this approach, all scores equal to or below 24 months were categorized in the one-year MA band. AE scores of 25 to 35 months were categorized in the two-year MA band.

2.3.3 Garage Game Task—Participants were administered the *Garage Game*, a measure of working memory (Devine et al., 2019). The *Garage Game* is an adapted version of the original *Three Boxes* task, a self-ordered pointing task devised to test working memory in humans and animal models (Diamond et al., 1997; Petrides, 1995). In the original *Three Boxes* task used with toddlers and young children, participants were presented with three containers with distinct colors and shapes on the lids, each containing a reward (Diamond et al., 1997). To obtain the rewards, participants were required to recall their previous container selections and avoid repeating selections. Between trials, the containers were hidden by a screen and a delay was imposed. When the boxes were revealed, participants were encouraged to find a reward by selecting and opening a container. Two versions of *Three Boxes* were administered by Diamond et al., one including location scrambling and one where the boxes remained stationary between trials. During each trial, participants were required to select a box, open it, and retrieve the reward. Toddlers were administered the task with three boxes, and young children were administered the task with six boxes.

In the *Garage Game* adaptation of the *Three Boxes* task (Devine et al., 2019), there are fewer motor demands and language confounds imposed, areas of relative challenge for many children with DS (Fidler, 2005). In the current *Garage Game* task, children were asked to find toy cars hidden in toy garages with matching-colored doors in up to three sets of trials (See photo of materials in Figure 1). Participants were not required to open and retrieve the toy, but rather, locations were selected via gesture (pointing or pointing approximations). The present version of this task also involved minimal receptive language demands during the instructions and no expressive language demands for participant responses. A gradual increase in working memory demands was included across three sets of trials. Before administering the task, a practice set was administered. If a participant was successful during the practice set, they continued to the first task set. In the first and second sets, three search locations were used and a distractor was imposed between trials by placing a screen over the search locations. In the third set, six search locations were used and a screen was imposed between trials. A stop rule was implemented so that the task was discontinued if a participant selected three consecutive empty garages during any set. If participants found all the cars in a set without reaching the stop rule, they moved on to the next set.

Practice Set: Three garage set without distractor.: In the Practice Set, the examiner introduced the child to three toy cars (i.e., red, green, and blue plastic cars, "Battat Toys 3-Car Garage") and encouraged the child to briefly interact with them. The garage toy was placed in front of the child just out of reach. After the child was familiar with the toy cars, the examiner said, "Let's park my cars!" and placed each car into the corresponding-colored garage with the child watching. The examiner then closed the three garage doors simultaneously, and asked the child, "Where's a car?" or "Find a car.". For the first selection, any garage chosen by the child contained a car. The examiner opened the garage that was indicated by the child and briefly praised the child, then placed the car nearby within the child's view. The examiner pointed to the now empty garage and stated, "Look, this garage is empty." The examiner then closed the garage door and asked the child to find another car. If the child chose the same garage as before, the examiner opened the garage and said, "This garage is empty. Let's try again!" If the child chose a new garage, the examiner said, "Great job!" and then placed the car with the previously found car, in view and out of reach. The examiner then pointed to the open garage and said, "Look, this garage is empty" and closed the garage door. The examiner repeated these steps until the child found all three cars or made three consecutive incorrect searches (i.e., searching an empty garage). If the child successfully found all three cars, the examiner opened all the garage doors and let the child briefly play with the toy cars.

Sets 1 and 2: Three garage set with distractor.: To begin Sets 1 and 2 of the *Garage Game*, the examiner started by saying, "You found all my cars! Let's park my cars" following the practice set. The examiner then introduced a distractor by holding an opaque poster board between the child and the garages prior to asking the child to find a car. The examiner removed the poster board and allowed the child to retrieve a car by pointing to a garage door. The examiner opened the garage that was indicated by the child and briefly praised the child, then said, "Good working! Now this car has to drive away" (could also say the car had to go to work, school, the park, etc.). The examiner showed the child that the car was going out of view for the child. The examiner pointed to the empty garage and stated, "Look, this garage is empty" and closed the garage door. For each subsequent trial, the examiner placed the poster board in front of the garage prior to asking the child to find a car and then removed the poster board. If the child chose a garage that had been previously chosen, the examiner opened the garage and said, Oh, it's not there. Let's try again!" If the child chose a new garage, the examiner said, "Good working" then stated that the car had to drive away and placed the car out of sight. The examiner repeated these steps until the child found all three cars or made three consecutive errors. If the child successfully found all three cars, the examiner opened all the garage doors and let the child briefly play with the toy cars.

Set 3: Six garage set with distractor.: The examiner introduced Set 3 by saying, "Look! We have more cars!" The examiner introduced an additional set of three toy cars (i.e., red fire engine, white ambulance, and blue police wooden cars, "Melissa & Doug Keys & Cars Rescue Garage") and a matching three-bay garage. The child was allowed to explore the new cars for about 30 seconds, then the examiner said, "Let's park all of my cars!" The examiner parked each car in the correctly colored garage and shut the doors simultaneously. When the second garage was introduced, the examiner ensured the red garage doors from the two garages were not positioned next to each other. The examiner then followed the steps from Set 2, but with six garages for the child to choose from.

2.3.3.1 Scoring.: Correct and incorrect search scores were created by totaling the number of novel (correct) search locations and the number of repeated (incorrect) search locations chosen by the participant. To account for some participants reaching a stop rule in set 1 or 2 and discontinuing the task, a "repetitive search" rate was calculated by dividing the total number of incorrect searches by the total number of cars the child had the opportunity to search for. Repeated searches in the same location were interpreted as indicating a lack of working memory maintenance. For example, if a child attempted all three sets, they had the opportunity to search for 12 cars in total. If that child repeated a search location twice, their repetitive search rate would be 2/12, or 0.167. A score above 1 was possible if a participant repeated a search location more times than the number of cars presented. For example, a participant could find a car during their first search, then select the same garage two more times, then search in a new location and find one more car, and then search previous locations three times in a row to reach a stop rule. The repetitive rate in this case would be 5/3, or 1.67. Repetitive search rate scores were binned into the following categories: (0) No errors – repetitive search rate = 0, (1) Some errors – repetitive search rate between 0 and 1, (2) Many errors – repetitive search rate > 1, and (3) Task not attempted [participant did not pass practice set or refused to continue].

A lower limit, or "floor", was defined as a participant seeming to engage meaningfully in the task but being unable to pass the practice set. Examiners also evaluated participant attention and opposition toward the task upon completion. Attention was defined as the participant attending to the examiner's instructions (e.g., eye gaze directed toward examiner and/or garages, body positioned toward the garages) and actively engaged for most of the task (e.g., pointing or approximating a point when prompted to find a car). Opposition was defined as the participant demonstrating resistance (e.g., pushing the garages away from them), attempting to evade the task or examiner (e.g., turning body away from the task), or causing a disruption to the task (e.g., removing task materials from the table).

2.4 Analyses

Descriptive analyses were used to assess the feasibility of the Garage Game by CA and MA. Feasibility was defined *a priori* as a threshold of 80% of participants who could successfully pass the practice set by locating all three cars. A benchmark of 80% for feasibility has been used in prior work evaluating outcome measures with children with DS and intellectual disabilities (e.g., Hessl et al., 2016; Schworer et al., 2022).

Floor effects were assessed by evaluating the number of participants who could not pass the practice set, and ceiling effects were determined by the number of participants who completed set 3 without repeating any search locations. Alluvial plots, a data visualization approach to demonstrate the association between categorical variables, were generated to visualize repetition errors across CA and MA groups. The relation between CA and MA and repeated searches across all trials sets and within each trial set was formally tested using negative binomial regressions (cf. Hilbe, 2011) to account for the skew and overdispersion of the outcome variable as a count measure.

Test-retest reliability was assessed using intraclass correlation coefficients (ICCs) for the subsample of children who completed Waves 1 and 2, two weeks apart. Criteria was set *a priori* as values > 0.75 indicating good reliability, and values .50 to .75 indicating moderate reliability (Koo & Li, 2016; Schworer et al., 2023). Paired samples *t*-tests were conducted to evaluate the potential presence of practice effects between Waves 1 and 2.

To determine if children utilized a specific spatially structured search pattern, such as selecting garages in a sequential order from left to right or right to left, a rater systematically coded each set to identify instances where a spatial search strategy was employed. Additionally, a second rater independently coded 20% of the participants, and a perfect agreement rate of 100% was achieved between the two raters. The percentage of participants who employed a search strategy on any of the sets, as well as on all the sets, was calculated. Logistic regression analyses were conducted to determine if the use of a search strategy was associated with CA or MA.

3. Results

To ensure there were no systematic cross-site differences in child performances, the main outcome measures (total correct searches, total incorrect searches, and repeated search rate) were compared by site using One-Way ANOVAs. Results indicated that there were no significant differences based on site for the three dependent variables (see Table 2 for results). There was also no evidence that total correct searches, incorrect searches, or repeated search rate differed based on sex, race, ethnicity, or type of DS (see Table 2 for results).

3.1 Feasibility

Of the 78 participants, nine (11.5%) were unable to pass the practice set and did not proceed to Set 1 of the *Garage Game*. All nine children who did failed to pass the practice set had a MA under 24 months. The CA of the children who did not pass the practice set ranged from 2 to 5 years. One additional child did not continue to the first test set due to behavioral refusal.

Of the 68 children who continued to the first test set, a majority (89.7%) successfully located all three toy cars and 70.6% did not repeat any previous search locations. Seven children repeated a search location three times in a row and, therefore, did not continue to the second test set. Set 2 was skipped for one child due to the child's inattention. Out of the 60 children who attempted the second test set, 95% found all three cars and 65% did not repeat any search locations. Three children repeated a search location three times in a row and did not continue to the third test set. One child refused to continue to the third set. Fifty-seven children attempted the third and final set. Approximately half of those children (49.1%) located all six cars and 24.6% did not repeat any selections.

In total, three participants were missing some or all data on this task due to refusal or inattention. All three of these participants had an MA under 24 months (at the floor of the SB-5). The CA of the children with missing data due to behavior ranged from 45 months to 63 months. Altogether, 19 children (24.4%) reached a stop rule before the start of the third set. The MA of children who reached a stop rule ranged from 13 months to 25 months. The CA of children who reached a stop rule ranged from 2 to 5 years, with 73.7% being 4 or younger. See Table 3 for a summary of participant completion on each set. The majority (92.3%) of participants demonstrated attention to the task. Opposition (e.g., resistance toward the task, attempts to evade, causing a disruption) was demonstrated by 11.5% of participants. Opposition was highest amongst children in the 2-year-old CA age range (50%). Across ages 3–8 years, oppositional behavior ranged from 0% to 17.6%. Overall, this measure was feasible for administration in this age range, and a range of scores and performances was observable, despite the infrequent presence of potentially interfering behaviors.

3.2 Performance and Developmental Sensitivity

Alluvial plots were generated to visualize repetitive searching across CA and MA groups (See Figures 2 and 3). The plots demonstrate how participants in each CA and MA group were stratified across the various repetitive search rate bins. Repetitive search rates ranged from 0 to 1.67. Twelve children were in the "no errors" bin (15.2%). Forty-nine children made "some errors" (62.8%). Seven children made "many errors" (9%). Ten children did not attempt the trial sets due to refusal or inability to pass the practice set (12.8%). The MA of the children in the "many errors" bin ranged from 18 months to 25 months, and the CA ranged from 36 months to 66 months.

Most participants who made a high number of repetitive errors had a CA of 3 years or less, with fewer children aged 4–5 making a high number of errors, and no children older than 5 years making a high number of errors. Likewise, the participants who did not attempt the task (due to refusal or failure on the practice set) were in the younger CA categories of 2–5 years. There were no participants with a CA of 2 or 3 years who performed without error. Children who made some errors were distributed across all CA bands, demonstrating emerging working memory skill across all involved CA categories.

The partitioning of repetition error bins was more defined across MA groups compared to CA groups. All participants in the “not attempted” bin had a MA of less than 2 years, and most participants with a high number of errors had a MA of less than 2 years. The remaining participants with a high repetitive search rate had an MA of 2 years. The participants who made some search errors were distributed across MA groups, maintaining the implication that working memory skills are emergent across MA groups. However, the percentage of participants who made no errors increased as MA increased.

To model the rate of repetitive searches as a function of CA and MA, a negative binomial regression was fit with the total count of repeated searches as an outcome and an offset for total opportunities. Working memory performance was strongly related to CA in years, $\exp(\beta) = 0.76$, se = 0.07, $p < .0001$ (95% CI: 0.65, 0.87). The rate of repetitive searches was estimated to be, on average, 24.5% lower with each 1-year increase in CA. The rate of repetitive searches was not significantly related to MA in years, $\exp(\beta) = 1.00$, se = 0.001, $p = 0.47$ (95% CI: 0.996, 1.00). CA also significantly predicted repetitive searches in the first and third trial sets $\exp(\beta) = 0.47$, se = 0.19, $p < .0001$ (95% CI: 0.31, 0.67) and $\exp(\beta) = 0.83$, se = 0.07, $p = .005$ (95% CI: 0.72, 0.94), respectively. However, CA did not predict repetitive searches in the second trial set ($\exp(\beta) = 1.02$, se = 0.17, $p = .88$). Working memory performance by CA can be seen in Table 4. The range of correct searches and repetitive search rates by CA and MA can be seen in Table 5.

3.3 Test-Retest Reliability and Practice Effects

Preliminary test-retest reliability was investigated using a sub-sample of those participants ($n = 27$) who returned for a second visit two weeks after the first assessment. Twenty-two of these children passed the practice set for both visits, and the scores of this subset were examined for test-retest reliability and practice effects. The observed ICCs for repetitive search rate and correct searches were 0.60 and 0.14 respectively, indicating moderate reliability for the main outcome of interest, but poor reliability for the correct search metric. No meaningful practice effects were observed from Wave 1 to Wave 2 (Table 6 below). Ten of the children who returned for a second visit were 2–5 years old, and twelve were 6–8 years old.

3.4 Utilization of a Search Strategy

A potential alternative explanation for performances could be the use of a spatial strategy on this task (moving in one direction across garage doors) that would not necessarily involve the reliance upon working memory. Among the participants, 69.12% used a spatially structured search strategy on at least one of the three sets. However, when considering the subset of 57 participants who completed all three sets, only 5 individuals (8.77%) consistently employed a spatial search strategy across all three sets. The use of a search strategy on at least one set was related to CA (See search strategy by set and CA in Figure 4). Specifically, each additional year of CA was associated with an average of 67% higher odds of using a spatial strategy at least once, $\exp(\beta) = 1.67$ (95% CI: 1.13, 2.63). However, no association was found between MA and the adoption of a spatial search strategy, $\exp(\beta) = 1.00$ (95% CI: 1.00, 1.01).

4. Discussion

The study of EF in young children with DS is critical because of the relevance of EF skills for academic achievement, daily living skills, and later employment outcomes (Tomaszewski et al., 2018; Will et al., 2017). Current accounts of the early development of EF in children with DS are limited because of a lack of laboratory measures appropriate for young ages and developmental ranges that are sensitive to detect small changes in the efficiency and performance of individuals across repeated measures. A lack of appropriate EF measures also prevents the accurate evaluation of much-needed potential treatments and interventions for young children with DS. As a part of the larger effort to establish valid and reliable outcome measures for DS research, this study evaluated a performance measure of working memory for use with young children with DS. The performances of 78 children with DS aged 2.5–8.67 years on an adapted version of a self-ordered pointing task ("the *Garage Game*") were analyzed, the psychometric properties examined, and both were, for the most part, acceptable. High levels of feasibility, scalability, and developmental sensitivity were observed within this CA range in DS. Minimal practice effects and floor effects were observed.

4.1 Feasibility

A promising finding of this evaluation of the *Garage Game* was its high degree of feasibility within the DS sample for the entire CA and MA range. Only 11.5% of participants were unable to pass the practice set. This clears the *a priori* threshold of 80% feasibility for this task in this age range and suggests that floor effects for administration would be within reason in potential treatment studies. Of those children who passed the practice set, most were able to locate all three cars in sets 1 and 2, which included a distractor, and very few children discontinued the task due to refusal or inattention. A particularly notable finding is the feasibility of this measure for use with even very young developmental levels (the one-year MA band). Many participants who passed the practice set ($n = 29$) had MAs within the one-year MA band. This finding demonstrates that this measure can be used to detect early emerging working memory foundations within the population of young children with DS. However, it is also of note that all participants who failed to pass the practice set had MAs below 24 months, which should inform the selection of this measure for a child within this CA range, but with an MA within the one-year age band. For early treatment studies that target working memory in those with one-year MA equivalent scores, it may be necessary to assess performances with the practice set to determine whether this is an appropriate outcome measure.

4.2 Scalability

In addition to the high degree of feasibility, the *Garage Game* also proved to be scalable (mastery level) within this sample. This measure was able to capture performances at the very low end of the continuum for performances, as 8 children who passed the practice set did not locate all three cars in set 1. Seven of these participants repeated a search location three times in a row, reaching a stop rule, demonstrating the possibility of capturing challenges with working memory in the context of a highly feasible task. The scalability is also notable in that, while a few participants who reached a stop rule had MAs under 24 months, many more participants with an MA under 24 months were able to successfully locate the three cars in set 1 without reaching the stop rule.

Additional evidence of scalability can be observed in the first, second, and third sets of the *Garage Game*. The task increases in difficulty by adding a distractor (sets 1 and 2) and using six, rather than three, garages (set 3). A small number of participants who continued to set 2 did not locate all three cars in that set, all of whom had MAs under 3 years. Out of 78 total participants, 57 continued to set 3 of the *Garage Game*, of which approximately half located all six cars. These 28 participants were represented across all MA bands. Overall, this pattern of performance demonstrates that this measure can be appropriately administered to children with DS within this MA and CA range. Also, variability in performance can be captured, particularly when examining the different levels of difficulty embedded within the three sets of the *Garage Game* task.

Scalability was also observed within the key dimension of interest: rate of repeated searches. Approximately half or more of the participants located all three cars in the first two sets without making any search errors (60.3% and 50% of 78, respectively). However, only a few participants were able to locate six cars in set 3 without any errors (17.9% of 78). Of those who attempted set 3, the percentage who made no search errors was still relatively low (24.6% of 57). This repetitive search rate outcome, which signifies the degree to which participants maintained their representations of where they had and had not searched, appears to be a scalable metric of working memory that is likely to detect incremental gains in this critical aspect of EF.

4.3 Developmental Sensitivity

The *Garage Game* also appeared to demonstrate adequate developmental sensitivity in terms of child CA. Repetitive search rates for the full task were significantly related to CA, and CA significantly predicted repetitive search rates for the first and third sets (but not for the second). However, it is notable that repetitive search rates were not associated with child MA. This is likely due to the narrower range of MA values compared to CA values, and the small number of participants with MA values of 4 and 5. The null results may also suggest an experiential element to the development of working memory related to CA, but not necessarily related to gains in other aspects of cognitive development. This finding aligns with the larger literature on working memory in DS, which has been found to be an area of vulnerability for the majority of participants with DS across many different studies, regardless of overall developmental status (Daunhauer et al., 2014, 2017; Loveall et al., 2017; Tungate & Conners, 2021; Van Deusen et al., 2022). Thus, the findings from this study reflect the broader challenges with working memory in this population, which seem to persist regardless of increases in MA. Future work should continue to examine the associations between CA, MA, and aspects of working memory in DS.

4.4 Practice Effects and Test-Retest Reliability

This study also yields preliminary evidence for test-retest reliability and a lack of practice effects in a subsample of participants. There were 27 participants from the overall sample who completed the same assessment battery, including the *Garage Game* task, two weeks after the original administration. Paired *t*-test analyses demonstrated no significant practice effects across these two administrations, although the magnitude of the mean repetitive search rate performance was modestly reduced at Wave 2 (from .30 to .24). Though the ICC for correct searches was found to be poor for Wave 1 and Wave 2, the ICC at the two data waves provides suggestive evidence of test-retest reliability for the repeated search rates, though the magnitude of association was somewhat lower than standards for test-retest reliability in the recent literature on psychometric properties for outcome measures in DS. These findings should be preliminary only, however. Though two-week time windows were the goal for participants, the implementation of this study coincided with the onset of the COVID-19 pandemic and many families who chose to participate when safety protocols allowed opted for only one wave of visits rather than two. As a result, only approximately one-third of the participants were included in test-retest analyses. Thus, findings should be interpreted with caution and future studies should seek to evaluate test-retest reliability with a larger sample.

4.5 Recommendations

Overall, results indicate that the *Garage Game* is a feasible laboratory task for measuring working memory in young children with Down syndrome. Although some participants in the younger CA and MA ranges were not able to complete the task, other participants at those ages were able to complete all three sets of the task. The task is, therefore, feasible for children as young as 30 months chronologically and under 24 months developmentally. Calculating a repetitive search rate for this task conveyed additional information about emerging working memory skills in young children with DS: the rate of repetitive search errors. The oldest participant in the sample (8 years chronologically) completed all three test sets without making any errors, as did a small number of participants at younger CAs. The majority of children, however, demonstrated some degree of repetitive search behavior. Based on results from this study, the *Garage Game* is recommended for measuring working memory performance in children with DS ages 2.5–7.99 years, with some caution needed for children who have MAs under 24 months.

4.6 Working Memory and the *Garage Game*

Our discussion would be incomplete without a reflection on the aspects of working memory that may be captured by the *Garage Game*. One important aspect of working memory is updating, or the ability to continuously revise information held in memory through the addition and deletion of content. The original *Three Boxes* task has been described by some as an updating task due to its requirement that participants remember and continuously update which locations have already been searched, and which locations remain unsearched (Völter et al., 2019). Other researchers have emphasized an inhibition aspect to this task by arguing that it requires that participants ignore competing stimuli and inhibit a prepotent response to return to a previously rewarded location (Devine et al., 2019). Still others simply describe this task as capturing working memory more broadly (Jenkins & Berthier, 2014). While we have taken the broader approach to describing the nature of the *Garage Game* as a working memory task, we should also consider that unique constructs like inhibition and updating may ultimately be driving performances as well.

4.7 Limitations and Future Directions

There were some limitations of this study. Because of the CA and MA range for this study, two measures were used to derive MA scores. Most participants over the age of 3 years were administered the SB-5, but out of concern for the MA floor of 24 months, many of the younger participants were also administered the Bayley-4 Cognitive domain. This allowed for those participants with scores at the SB-5 floor to have values that better approximated their overall cognitive status. However, not all participants were administered both assessments for time-related reasons, and as a result, some participants did receive MA scores at the floor for the SB-5. Ideally, one measure would have been available for the entire sample of participants, but to date, no such measure exists to assess cognitive status across this chronological and developmental age range. The Bayley-4 can determine AE lower than the SB-5 and would have been a more suitable assessment of developmental status for all who scored at the floor of the SB-5. However, the Bayley-4 is normed only up to the age of 42 months, and therefore, would not have been a valid assessment for chronologically older participants included in the study. This difficulty was addressed by conducting analyses using one-year age bands with CA and MA, and all participants at the 24-month floor for the SB-5 were assigned to the one-year MA band. Despite this issue, MA-related findings should be interpreted with some degree of caution due to the use of two measures to derive age equivalent scores, and the presence of some participants at the floor of the SB-5. Furthermore, age-equivalent scores such as those generated by the Bayley-4 and SB-5 represent the median level of performance of children from the norming sample at that chronological age (Conrad, 2018) and are not a perfect assessment of developmental status. Therefore, measures that construct norm references specific to children with DS, such as the *Garage Game*, are needed.

An additional aspect that warrants attention is the performance distribution observed among the participants. It is important to note that nine children failed to pass the practice set, while 14 participants achieved perfect scores on set 3. This indicates that approximately 29% of the sample (23 out of 78 participants) exhibited either floor or ceiling effects. The presence of floor and ceiling effects may introduce limitations to the interpretation of the results, as they suggest potential constraints regarding the sensitivity and range of the measurement.

Furthermore, this study was not designed to establish the construct validity of this measure through convergence with other accepted measures of working memory. This absence is an inherent consequence of the need we have highlighted for suitable measures of EF designed for young children with DS. As additional measures undergo evaluation, this will be a priority for future investigation. It is noted that this task exhibits face validity, which can serve as a starting point for this future work.

In the future, the *Garage Game* should be evaluated for its performance over longer periods of time, to determine if the task is appropriate for measuring change in the context of clinical trials or intervention. Examining the performance of more participants in higher CA and MA age bands will be an important next step for future research. Future work should also seek to establish alternate measures of working memory that are feasible with young participants with DS. Finally, future work may explore the adaptability of the *Garage Game* to other contexts. For example, the task could potentially be administered as a tablet- or computerbased assessment, similar to other existing measures of visuo-spatial memory including the Cambridge Neuropsychological Test Automated Battery Paired Associates Learning task (CANTAB PAL; https://cambridgecognition.com/paired-associates-learning-pal/).

5. Conclusions

Overall, there is preliminary evidence for the utility of a modified self-ordered pointing task, the *Garage Game*, for use with children with DS in potential treatment studies. The measure was found to capture adequate variability in performance and was developmentally sensitive within this CA range. Preliminary evidence for modest test-retest reliability and minimal practice effects suggest that this measure can capture skill acquisition in the critical dimension of working memory. These findings contribute to the growing knowledge base regarding the assessment of developmental outcomes in children with DS, who have been historically under-represented in treatment studies in the field of developmental disabilities research. With a more advanced understanding of outcome measurement in this population, more advanced study designs can be implemented, with the possibility of detecting treatment effects with accuracy and precision.

Funding Acknowledgements

Research reported in this article was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award numbers R21 HD101000, R01HD099150, and P50HD103526.

References

Baddeley AD (1983). Working memory. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 302(1110), 311–324.

Bayley N, & Aylward GP (2019). Bayley Scales of Infant and Toddler Development fourth edition (Bayley-4). Bloomington, MN: NCS Pearson.

Bennett SJ, Holmes J, & Buckley S (2013). Computerized memory training leads to sustained improvement in visuospatial short-term memory skills in children with Down syndrome. American Journal on Intellectual and Developmental Disabilities, 118(3), 179–192. [PubMed: 23734613]

Best JR, & Miller PH (2010). A developmental perspective on executive function. Child Development, 81(6), 1641–1660. 10.1111/j.1467-8624.2010.01499.x [PubMed: 21077853]

Blair C, Zelazo PD, & Greenberg MT (2005). The measurement of executive function in early childhood. In Developmental Neuropsychology (Vol. 28, Issue 2, pp. 561–571). 10.1207/s15326942dn2802_1 [PubMed: 16144427]

Conners FA, Rosenquist CJ, Arnett L, Moore MS, & Hume LE (2008). Improving memory span in children with Down syndrome. Journal of Intellectual Disability Research, 52(3), 244–255. [PubMed: 18261023]

Conrad Z (2018). Age-equivalent scores. In Frey BB (Ed.), The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation (p. 62). Sage Publications.

Danielsson H, Henry L, Rönnberg J, & Nilsson L-G (2010). Executive functions in individuals with intellectual disability. Research in Developmental Disabilities, 31(6), 1299–1304. [PubMed: 20728303]

Daunhauer LA, & Fidler DJ (2013). Executive functioning in individuals with Down syndrome. In Handbook of Self-Regulatory Processes in Development (pp. 458–477). Psychology Press. 10.4324/9780203080719-32

Daunhauer LA, Fidler DJ, Hahn L, Will E, Lee NR, & Hepburn S (2014). Profiles of everyday executive functioning in young children with Down syndrome. American Journal on Intellectual and Developmental Disabilities, 119(4), 303–318. 10.1352/1944-7558-119.4.303 [PubMed: 25007296]

Daunhauer LA, Gerlach-McDonald B, Will E, & Fidler DJ (2017). Performance and ratings based measures of executive function in school-aged children with Down syndrome. Developmental Neuropsychology, 42(6), 351–368. 10.1080/87565641.2017.1360303 [PubMed: 28985480]

Devine RT, Ribner A, & Hughes C (2019). Measuring and predicting individual differences in executive functions at 14 months: A longitudinal study. Child Development, 90(5), e618–e636. 10.1111/cdev.13217 [PubMed: 30663776]

Diamond A, Barnett WS, Thomas J, & Munro S (2007). The early years: Preschool program improves cognitive control. Science, 318(5855), 1387–1388. 10.1126/science.1151148 [PubMed: 18048670]

Diamond A, Prevor MB, Callender G, & Druin DP (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. 62(4), 1–206. https://www.jstor.org/stable/1166208

Edgin JO, Anand P, Rosser T, Pierpont EI, Figueroa C, Hamilton D, Huddleston L, Mason G, Spaǹo G, Toole L, Nguyen-Driver M, Capone G, Abbeduto L, Maslen C, Reeves RH, & Sherman S (2017). The Arizona cognitive test battery for down syndrome: Test-retest reliability & practice effects. American Journal on Intellectual and Developmental Disabilities, 122(3), 215–234. 10.1352/1944-7558-122.3.215 [PubMed: 28452581]

Edgin JO, Clark C, Massand E, & Karmiloff-Smith A (2015). Building an adaptive brain across development: Targets for neurorehabilitation must begin in infancy. Frontiers in Behavioral Neuroscience, 9. 10.3389/fnbeh.2015.00232

Elliott CD (2007). Differential ability scales (2nd ed.). Harcourt Assessment.

Esbensen AJ, Hooper SR, Fidler D, Hartley SL, Edgin J, D'Ardhuy XL, Capone G, Conners FA, Mervis CB, Abbeduto L, Rafii M, Krinsky-Mchale SJ, Urv T, Dykens E, Esbenson A, Hartlay S, Keller S, & Weir S (2017). Outcome measures for clinical trials in down syndrome. American Journal on Intellectual and Developmental Disabilities, 122(3), 247–281. 10.1352/1944-7558-122.3.247 [PubMed: 28452584]

Fidler DJ (2005). The emerging Down syndrome behavioral phenotype in early childhood. Infants & Young Children, 18(2), 86–103. 10.1097/00001163-200504000-00003

Garon N, Bryson SE, & Smith IM (2008). Executive function in preschoolers: A review using an integrative framework. Psychological Bulletin, 134(1), 31–60. 10.1037/0033-2909.134.1.31 [PubMed: 18193994]

Grieco J, Pulsifer M, Seligsohn K, Skotko B, & Schwartz A (2015). Down syndrome: Cognitive and behavioral functioning across the lifespan. American Journal of Medical Genetics, Part C: Seminars in Medical Genetics, 169(2), 135–149. 10.1002/ajmg.c.31439 [PubMed: 25989505]

Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, & Kirby J (2019). The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics, 95, 103208. [PubMed: 31078660]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42(2), 377–381. [PubMed: 18929686]

Hessl D, Sansone SM, Berry-Kravis E, Riley K, Widaman KF, Abbeduto L, Schneider A, Coleman J, Oaklander D, & Rhodes KC (2016). The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. Journal of Neurodevelopmental Disorders, 8(1), 1–18. 10.1186/s11689-016-9167-4 [PubMed: 26855682]

Hilbe JM (2011). Negative Binomial Regression (2nd ed.). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511973420

Isaacs EB, & Vargha-Khadem F (1989). Differential course of development of spatial and verbal memory span: A normative study. British Journal of Developmental Psychology, 7(4), 377–380. 10.1111/j.2044-835X.1989.tb00814.x

Jarrold C (2001). Applying the working memory model to the study of atypical development. In Andrade J (Ed.), Working Memory in Perspective (pp. 126–150). Psychology Press.

Jenkins IL, & Berthier NE (2014). Working memory and inhibitory control in visually guided manual search in toddlers. Developmental Psychobiology, 56(6), 1252–1262. 10.1002/dev.21205 [PubMed: 24752642]

Koo TK, & Li MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012 [PubMed: 27330520]

Lanfranchi S, Jerman O, Dal Pont E, Alberti A, & Vianello R (2010). Executive function in adolescents with Down Syndrome. Journal of Intellectual Disability Research, 54(4), 308–319. 10.1111/j.1365-2788.2010.01262.x [PubMed: 20202074]

Lanfranchi S, Pulina F, Carretti B, & Mammarella IC (2017). Training spatialsimultaneous working memory in individuals with Down syndrome. Research in Developmental Disabilities, 64, 118–129. 10.1016/j.ridd.2017.03.012 [PubMed: 28388504]

Loveall SJ, Conners FA, Tungate AS, Hahn LJ, & Osso TD (2017). A crosssectional analysis of executive function in Down syndrome from 2 to 35 years. Journal of Intellectual Disability Research, 61(9), 877–887. 10.1111/jir.12396 [PubMed: 28726285]

Mai CT, Isenburg JL, Canfield MA, Meyer RE, Correa A, Alverson CJ, Lupo PJ, Riehle-Colarusso T, Cho SJ, Aggarwal D, Kirby RS, & Network NBDP (2019). National population-based estimates for major birth defects, 2010–2014. Birth Defects Research, 111(18), 1420–1435. 10.1002/bdr2.1589 [PubMed: 31580536]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, & Wager TD (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49–100. 10.1006/cogp.1999.0734 [PubMed: 10945922]

Petrides M (1995). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. Journal of Neuroscience, 15(1), 359–375. [PubMed: 7823141]

Roid GH (2003a). Stanford Binet intelligence scales, technical manual (5th ed.). Riverside Publishing.

Roid GH (2003b). Stanford-Binet Intelligence Scales–Fifth Edition. Itasca, IL: Riverside Publishing.

Rowe J, Lavender A, & Turk V (2006). Cognitive executive function in Down’s syndrome. British Journal of Clinical Psychology, 45(1), 5–17. 10.1348/014466505X29594 [PubMed: 16480563]

Sansone SM, Schneider A, Bickel E, Berry-Kravis E, Prescott C, & Hessl D (2014). Improving IQ measurement in intellectual disabilities using true deviation from population norms. Journal of Neurodevelopmental Disorders, 6, 1–14. 10.1186/1866-1955-6-16 [PubMed: 24433325]

Schworer EK, Esbensen AJ, Fidler DJ, Beebe DW, Carle A, & Wiley S (2022). Evaluating working memory outcome measures for children with Down syndrome. Journal of Intellectual Disability Research, 66(1–2), 195–211. 10.1111/jir.12833 [PubMed: 33763953]

Schworer EK, Hoffman EK, & Esbensen AJ (2021). Psychometric evaluation of social cognition and behavior measures in children and adolescents with down syndrome. Brain Sciences, 11(7). 10.3390/brainsci11070836

Schworer EK, Soltani A, Altaye M, Fidler DJ, & Esbensen AJ (2023). Cognitive flexibility assessment in youth with Down syndrome: Reliability, practice effects, and validity. Research in Developmental Disabilities, 133, 104416. 10.1016/j.ridd.2022.104416 [PubMed: 36603310]

Smeyne CN, Esbensen AJ, Schworer EK, Belizaire S, Hoffman EK, Beebe DW, & Wiley S (2022). Evaluating verbal fluency outcome measures in children with Down syndrome. American Journal on Intellectual and Developmental Disabilities, 127(4), 328–344. [PubMed: 36122330]

Swanson HL, & Alloway TP (2012). Working memory, learning, and academic achievement. In APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues (pp. 327–366). American Psychological Association. 10.1037/13273-012

Tomaszewski B, Fidler D, Talapatra D, & Riley K (2018). Adaptive behaviour, executive function and employment in adults with Down syndrome. Journal of Intellectual Disability Research, 62(1), 41–52. 10.1111/jir.12450 [PubMed: 29214700]

Toplak ME, West RF, & Stanovich KE (2013). Practitioner review: Do performancebased measures and ratings of executive function assess the same construct? Journal of Child Psychology and Psychiatry and Allied Disciplines, 54(2), 131–143. 10.1111/jcpp.12001 [PubMed: 23057693]

Tungate AS, & Conners FA (2021). Executive function in Down syndrome: A metaanalysis. Research in Developmental Disabilities, 108, 103802. 10.1016/j.ridd.2020.103802 [PubMed: 33341075]

van Deusen K, Prince MA, Esbensen AJ, Edgin JO, Schworer EK, Thurman AJ, Patel LR, Daunhauer LA, & Fidler DJ (2022). Profiles of caregiver-reported executive function in children with Down syndrome. Brain Sciences, 12(10). 10.3390/brainsci12101333

Völter CJ, Mundry R, Call J, & Seed AM (2019). Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task. Proceedings of the Royal Society B: Biological Sciences, 286(1907). 10.1098/rspb.2019.0715

Wechsler D (2014). Wechsler Intelligence Scale for Children (5th ed.). Pearson.

Will E, Fidler DJ, Daunhauer L, & Gerlach-McDonald B (2017). Executive function and academic achievement in primary - grade students with Down syndrome. Journal of Intellectual Disability Research, 61(2), 181–195. 10.1111/jir.12313 [PubMed: 27561217]

Zelazo PD, & Carlson SM (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. Child Development Perspectives, 6(4), 354–360. 10.1111/j.1750-8606.2012.00246.x

Highlights:

- The Garage Game task was feasible for children with Down syndrome, even those with young chronological and mental ages
- The task was scalable and able to capture a range of performance abilities
- Performance on the Garage Game task was associated with chronological age.
- The task demonstrated preliminary evidence of test-retest reliability.

What this paper adds?

Reliable and feasible outcome measures are critical for accurate assessment in research and clinical trials. Standardized and normed assessments have often not been tested for use in individuals with Down syndrome. Executive functioning is an important target for intervention and research in Down syndrome because of its relevance for daily life and achievement; however, few measures of executive function have been rigorously tested for use with individuals with Down syndrome, especially in younger chronological ages. This study adds to a growing body of literature that focuses on psychometrically evaluating measures of cognitive performance in individuals with Down syndrome. The study evaluates a measure of working memory in children ages 2–8 years old with Down syndrome for feasibility, reliability, and sensitivity, describes performance by subgroups of chronological age and mental age bins, and makes recommendations for the measure's use in potential future research and trials..

Figure 1.
Garage Game materials.

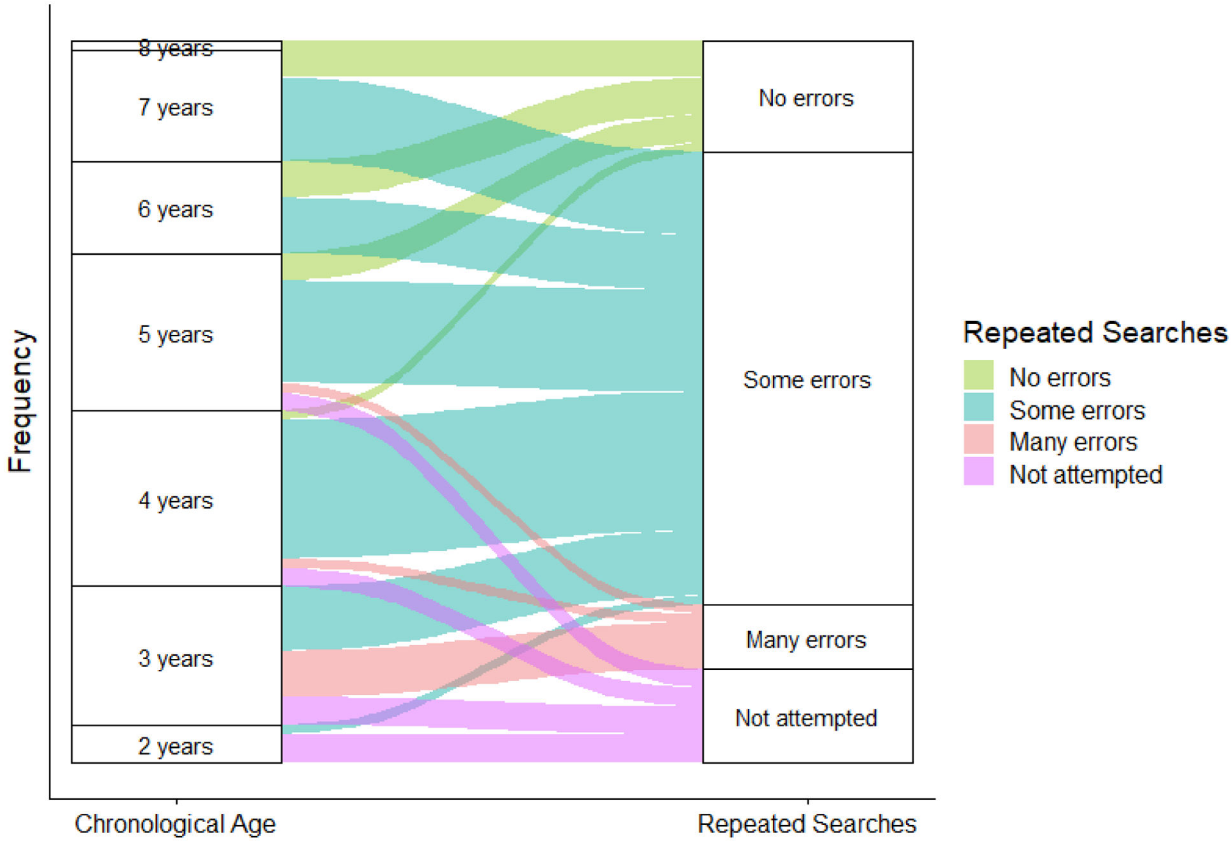

Figure 2.

Plot of repetitive search bins by CA.

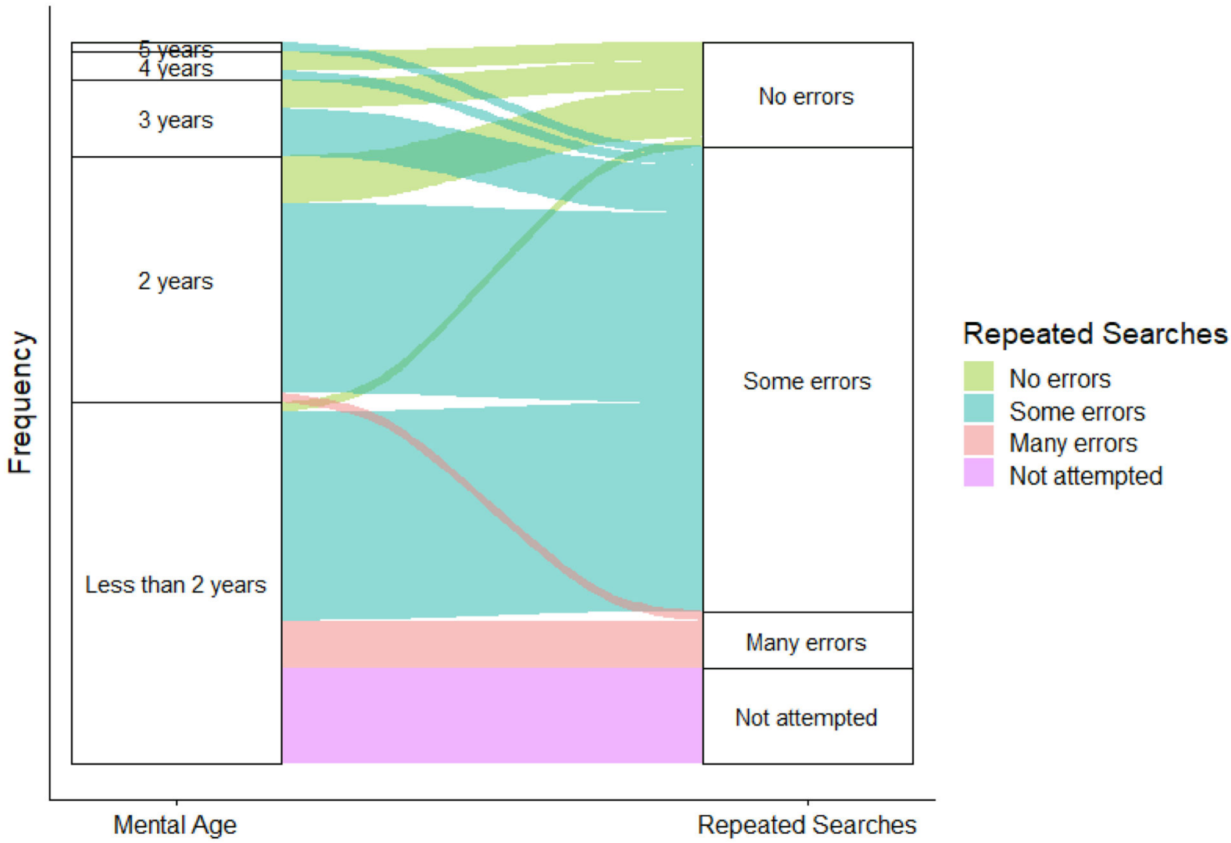

Figure 3.
Plot of repetitive search bins by MA.

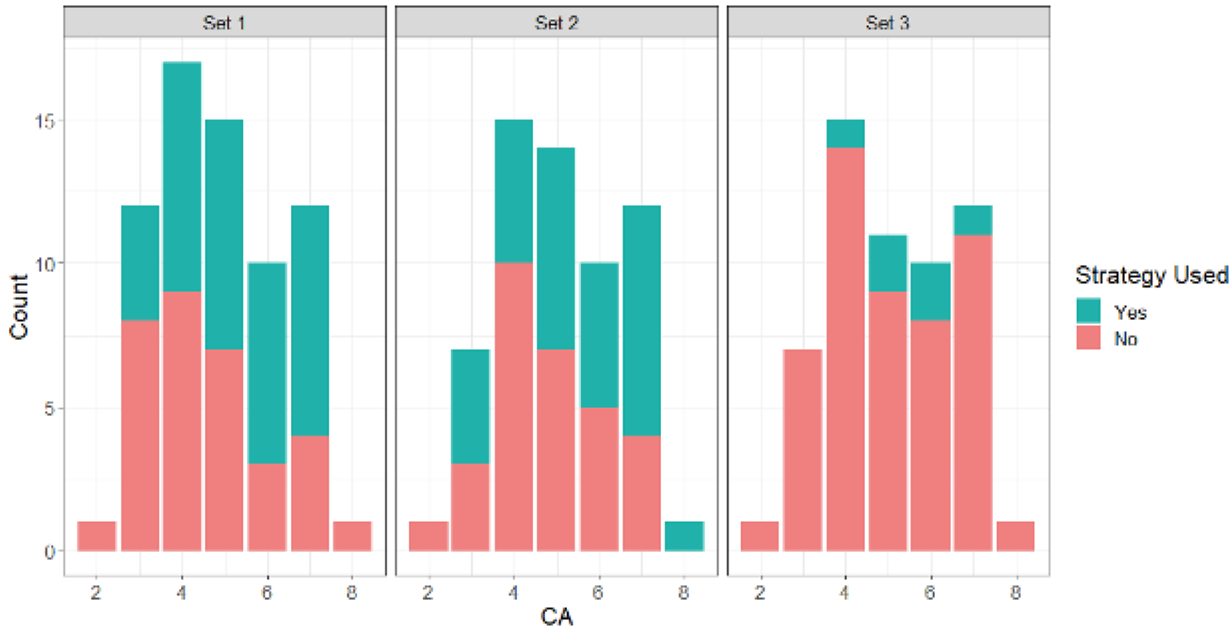

Figure 4.
Plot of spatial search strategy by set and CA.

Table 1.

Demographic information (n=78).

Child Characteristics	Mean (SD) or % (n)
Male	50% (39)
Chronological Age (years)	5.17 (1.49)
Race (n = 6 missing)	
Asian	3.8% (3)
Black / African American	2.6% (2)
White	78.2% (61)
Multiple / Other	6.4% (5)
Unknown	1.3% (1)
Ethnicity (n = 8 missing)	
Hispanic or Latino	12.8% (10)
Not Hispanic or Latino	76.9% (60)
DS Type (n = 2 missing)	
Trisomy 21	87.2% (68)
Mosaicism	1.3% (1)
Translocation	3.8% (3)
Not Sure	5.1% (4)
Premature Birth (% yes; n = 4 missing)	23.1% (18)
Congenital Heart Defects (% yes, n = 4 missing)	66.7% (52)
Autism Spectrum Disorder (% yes, n = 4 missing)	3.85% (3)
Attention Deficit and/or Hyperactivity Disorder (% yes, n = 4 missing)	1.28% (1)

Table 2.

Group comparisons ANOVA results.

	df	*F*	*p*
Total Correct Searches			
Site	3, 64	0.45	.72
Sex	1, 66	0.02	.89
Race	4, 57	0.07	.99
Ethnicity	1, 59	0.64	.43
DS Type	3, 62	0.93	.43
Total Incorrect Searches			
Site	3, 64	0.94	.43
Sex	1, 66	0.12	.74
Race	4, 57	1.18	.33
Ethnicity	1, 59	0.1	.75
DS Type	3, 62	0.66	.58
Repetitive Search Rate			
Site	3, 64	1.0	.4
Sex	1, 66	0.49	.49
Race	4, 57	0.32	.87
Ethnicity	1, 59	0.39	.53
DS Type	3, 62	0.55	.65

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Sample performance across sets of the *Garage Game* task.

	Practice Set	Set 1	Set 2	Set 3
Attempted	78	68	60	57
Completed	69 (89.7%)	61 (89.7%)	57 (95%)	28 (49.1%)
No repeated searches	-	47 (70.1%)	39 (65%)	14 (24.6%)
Missing	-	1 (refusal)	1 (inattention)	1 (refusal)

Table 4.

Repetitive search rate bins by chronological age.

Repetitive Rate Bin	**Chronological Age Bin % (n)**						
	2	**3**	**4**	**5**	**6**	**7**	**8**
No errors	0 (0)	0 (0)	5.9 (1)	20 (3)	40 (4)	25 (3)	100 (1)
Some errors	100 (1)	53.8 (7)	88.2 (15)	73.3 (11)	60 (6)	75 (9)	0 (0)
Many errors	0 (0)	38.5 (5)	5.9 (1)	6.7 (1)	0 (0)	0 (0)	0 (0)
Not attempted	0 (0)	7.7 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Table 5.

Range of task performance by CA and MA.

Total Correct Searches [min, max]		Repetitive Search Rate [min, max]	
MA	**CA**	**MA**	**CA**
1 [2, 12]	2 9	1 [0, 1.67]	2 0.58
2 [2, 12]	3 [2, 12]	2 [0, 1.00]	3 [0.08, 1.67]
3 [8, 12]	4 [2, 12]	3 [0, 0.58]	4 [0, 1.00]
4 12	5 [2, 12]	4 [0, 0.08]	5 [0, 1.00]
5 12	6 [8, 12]	5 0.33	6 [0, 0.67]
	7 [8, 12]		7 [0, 0.58]
	8 12		8 0

* Single number represents only one score that was in this CA or MA range.

Table 6.

Practice effects from Wave 1 to Wave 2.

	Wave 1 Mean (SD)	**Wave 2 Mean (SD)**	***t***	***p***	**Cohen's *d***
Repetitive Search Rate	0.30 (0.27)	0.24 (0.24)	1.10	.28	0.23
Total Correct Searches	10.27 (2.62)	10.86 (1.36)	−1.01	.32	2.74