languages

MDPI

Article

# Imitating the Robots: Measuring Memory Flexibility with Monolingual and Bilingual Preschoolers

**Joscelin Rocha-Hidalgo [1,*], Sylvia N. Rusnak [1], Olivia A. Blanchfield [1], Sharanya Suresh [2], Lily Tahmassebi [1], Hadley Greenwood [1], Kimberly Chanchavac [1] and Rachel Barr [1]**

[1] Department of Psychology, Georgetown University, Washington, DC 20057, USA
[2] College of Human Ecology, Cornell University, Ithaca, NY 14850, USA
* Correspondence: jr1679@georgetown.edu

**Abstract:** Millions of children in the United States are growing up hearing multiple languages. Memory flexibility is the ability to apply information from a past experience to future situations that are perceptually different from the initial learning experience and differs between monolinguals and bilinguals during infancy. We use a new, non-verbal object sequencing imitation task (OSI) to measure memory flexibility changes in monolingual and bilingual preschoolers. In the OSI task, children imitate target actions to produce a final pose on a robot figure. Children are tested with different robots than those used to demonstrate the target actions to test memory flexibility. We hypothesized that both monolingual and bilingual children would imitate the sequences significantly above baseline, but bilingual preschoolers would do so at a greater rate than their monolingual peers. To test this hypothesis, we visited 101 3-year-olds in their homes. An experimenter demonstrated 2- to 5-step sequences on one robot, and children were tested on a functionally similar but perceptually different robot. All preschoolers performed significantly above baseline on the total composite percentage score (the correct number of movements and pairs summed across all sequences, divided by the possible maximum score). There were no significant differences between monolinguals and bilinguals in baseline and test trials. We repeated the same pattern of results using a multi-level model, including all trials. The common binary classification of bilinguals and monolinguals often does not adequately describe the complex experience of growing up in a bilingual environment. Modeling the heterogeneity that arises from growing up in a bilingual home is important for understanding how this arrangement could impact an individual's cognitive development. To consider such heterogeneity, we implemented latent profile modeling to identify language groups based on a series of variables such as L2, L3 exposure, speakers' nativeness to the languages, and speakers' proficiency and identified three profiles (low, medium, and high multilingual exposure). The pattern of results remained the same. We conclude that memory flexibility differences exhibited during infancy may plateau during early childhood.

**Keywords:** preschoolers; memory load; imitation; memory flexibility; working memory



**Citation:** Rocha-Hidalgo, Joscelin, Sylvia N. Rusnak, Olivia A. Blanchfield, Sharanya Suresh, Lily Tahmassebi, Hadley Greenwood, Kimberly Chanchavac, and Rachel Barr. 2022. Imitating the Robots: Measuring Memory Flexibility with Monolingual and Bilingual Preschoolers. *Languages* 7: 268. https://doi.org/10.3390/languages7040268

Academic Editors: Leher Singh and Marina Kalashnikova

Received: 26 March 2022
Accepted: 11 October 2022
Published: 21 October 2022



## 1. Introduction

Growing up in a multilingual home is a common experience in the United States during early childhood. By 2019, approximately 12 million children were living in homes where a language other than English was spoken (Kids Count Data Center 2019). Researchers have become increasingly interested in the role that growing up in a multilingual home plays in executive functioning, particularly cognitive flexibility and working memory (see Adi-Japha et al. 2010; Barac et al. 2014). Working memory is the short-term ability to retain, manipulate, and update information (Baddeley and Hitch 1974). Although few studies have examined precursors of cognitive flexibility, one potential precursor is memory flexibility (MF).

Memory flexibility is the ability to apply information from a past experience to future situations that are perceptually different from the initial learning experience (Eichenbaum 1997; Hayne 2006; Karmiloff-Smith 1994). This process allows a child to generalize beyond the specific details of the original encoding context to apply what they learned to diverse problems (Barr and Brito 2014; Brito et al. 2019). Early in development, memory is contingent on an exact match between encoding and retrieval conditions. With age, children can increasingly tolerate differences between the encoding cues and those available at test (Brito et al. 2019; Brito and Barr 2014).

Imitation paradigms are a robust measure of memory development and have been used to assess memory flexibility differences between monolingual and bilingual infants and toddlers. Imitation improves with age and correlates across ages. In one study, infants were tested at 12, 24, and 36 months of age. There were age-related increases in imitation across time, and some performance stability, with cross-age correlations highest between 24 and 36 months (Rose et al. 2005). Over an even more extended period, immediate nonverbal recall on imitation tasks at 20 months of age was significantly associated with nonverbal memory at six years of age (Riggins et al. 2013). Performance on imitation tasks in infancy is also associated with later cognitive outcomes: using a median split approach, infants who performed poorly on 1-step imitation tasks at nine months of age had poorer general cognitive abilities as measured by the McCarthy Scales of Children's Abilities at four years of age than infants who had performed well (Strid et al. 2006). While previous tasks have been able to discriminate cross-age correlations in memory performance and associations with later cognitive outcomes (for review, see Brito et al. 2019), the restricted range of scores may have limited the predictive value of imitation learning. However, because other imitation tasks used with preschoolers were not explicitly designed to test memory flexibility (e.g., Williamson et al. 2010), the relationship between language exposure (monolingual vs. bilingual) and memory flexibility and working memory in preschoolers has not been examined.

Rusnak and colleagues (Rusnak et al. 2022) developed an imitation task to measure working memory and memory flexibility in preschoolers, the Object Sequencing Imitation (OSI). We used the OSI to measure memory flexibility in monolingual and bilingual preschoolers. In the OSI task, children imitate target actions to produce a final pose on a wooden robot figurine. Rusnak and colleagues developed 2- to 5-step sequences and tested 3- to 5-year-olds. They calculated baselines for each sequence because well-designed imitation tasks must have low (at or near zero) and age-invariant baselines (e.g., Barr and Hayne 2000; Meltzoff 1990) and found that baselines were uniformly low. They demonstrated that the experimental group performed significantly above the baseline control. Increasing the number of items to remember increases the working memory load. Consistent with other research (e.g., Barr et al. 2016), Rusnak and colleagues found that performance on the OSI task varied as a function of low versus higher cognitive load across multiple trials in 3- to 5-year-olds. The task was parameterized by age, and there were no age-related differences. With multiple trials and multiple sequences, successful imitation of the robot sequences requires that children update their memory from one pose to the next and measure visuospatial working memory. The children imitated the target actions on a novel wooden robot figurine exhibiting memory flexibility.

### *1.1. Memory Flexibility and Bilingualism*

Memory flexibility is measured using generalization imitation tasks, which require children to encode, retain, and retrieve a memory, even when perceptual details of the objects change between encoding and retrieval. Bilingual infants and toddlers demonstrate a different earlier trajectory of memory flexibility than monolinguals (e.g., Barr et al. 2019; Brito and Barr 2012, 2014; Brito et al. 2014, 2015). Six- (Brito and Barr 2014) and 18-month-old (Brito and Barr 2012; Brito et al. 2015) bilinguals performed above baseline on a puppet generalization task, demonstrating that at a very young age, bilingual infants can transfer the behaviors learned with puppet A onto a novel puppet B at a higher rate than their monolingual peers. Furthermore, 18- (Barr et al. 2019; Brito et al. 2021) and 24-month-old bilinguals (Brito et al. 2014, 2021) performed above baseline on the animal and rattle generalization task after a 24-h delay, while monolinguals did not. It is important to note that monolinguals and bilinguals did not differ on a memory recall version of the imitation task when the perceptual features did not change between encoding and retrieval (Barr et al. 2019; Brito et al. 2014) or on an immediate working memory task (Brito et al. 2014, 2021). Infants growing up in a bilingual environment are exposed to more varied speech patterns than monolingual infants and may have more practice using a wider range of retrieval cues (Brito et al. 2014). Despite the evidence that language exposure possibly alters children's memory flexibility skills, until recently, no suitable tasks have been developed for children above two. The OSI task generates a wide variability in scores and can be used to test preschoolers making it the ideal task to test whether there are memory flexibility and working memory differences between monolinguals and bilingual preschoolers.

### *1.2. The Present Study*

In an imitation paradigm, an experimenter demonstrates a series of actions on an object. Following a delay ranging from a few seconds to several weeks, children are given the opportunity to perform what they observed during the demonstration (Barr et al. 1996). Following earlier memory flexibility protocols (Brito and Barr 2012; Rusnak et al. 2022), children in the present study are tested with different robots than those used to demonstrate the target actions to test memory flexibility. Furthermore, the nonverbal nature of the OSI task ensures that preschoolers do not perform poorly because of verbal demands (Brito et al. 2019). Using the 2- to 5-step sequences developed by Rusnak and colleagues (Rusnak et al. 2022) and the multiple-sequence design allowed us to analyze children's imitation performance at each trial using multi-level models and create a composite percentage score. Rocha-Hidalgo and Barr (2022) conducted a scoping review to examine extant literature defining bilinguals in children under three years. They suggested that researchers consider not only binary Monolingual vs. Bilingual categorization but also consider using a continuous approach and latent profile analyses to categorize participants. We adopted these three approaches in the current study and examined whether the pattern of results converged. We hypothesized that both monolingual and bilingual children would imitate the sequences, indexed by performance using their composite scores across all poses and their trial-based imitation scores, significantly above baseline but that bilingual preschoolers would do so at a greater rate than their monolingual peers across trials.

## 2. Materials and Methods

The present study was conducted according to the guidelines in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at Georgetown University.

### *2.1. Participants*

A total of 101 (48 girls) 3-year-old children ($M_{age}$ = 39.47 months, $SD$ = 2.54) were recruited and tested in their homes between August 2016 and March 2022 as part of an ongoing longitudinal study (see Rocha-Hidalgo et al. 2021 for additional study from this project). Of the parents who provided demographic questions ($N$ = 98), 76 reported their child to be White/Caucasian, 4 Asian/Asian American, 1 African/African American, and 17 Mixed. Nineteen children were identified by their caregivers as Latino/a/x. Participants were primarily from college-educated families, with a mean number of 17.57 years of education ($SD$ = 1.35; averaged between parents). Most of the children were from middle- to high-income homes, with an average yearly income of $93,765.55 ($SD$ = $30,463.21) based on the median household income zip code for the family's postcode at the time of participation. Additional children were excluded from the analysis due to failure to interact with the experimental stimuli ($N$ = 5), experimenter error ($N$ = 3), and technology problems ($N$ = 7). Our criteria for failure to interact with stimuli include children who completed less than 50% of the test trials. There were no significant differences between Monolinguals and Bilinguals or among the profiles for the following variables: age, median household income, and average parental education.

All children in the final samples were exposed to English or Spanish. The dominant languages for children in the final sample for the Analysis Approach 1 were: English ($N$ = 63), Spanish ($N$ = 7), Mandarin ($N$ = 2), German ($N$ = 1), Korean ($N$ = 1), Russian ($N$ = 1), and Slovak ($N$ = 1). The dominant languages for children in the final sample for the Analysis Approach 2 were: English ($N$ = 84), Spanish ($N$ = 10), Mandarin ($N$ = 2), German ($N$ = 1), Korean ($N$ = 1), Russian ($N$ = 1), and Slovak ($N$ = 1).

### *2.2. Materials*

#### 2.2.1. NIH Toolbox Picture Vocabulary Test

The Picture Vocabulary Test was administered on an iPad (9.7-inch retina display) through the NIH toolbox application to capture children's receptive vocabulary. In each trial, children hear a target word and are shown four images on an iPad screen. The child then chooses the image that best matches the word heard. The toolbox calculates a theta score reflecting the child's receptive vocabulary score accounting for the difficulty of the item and the probability of the response being by chance (Item Response Theory; Gershon et al. 2014). The NIH Toolbox then calculates age-corrected scores based on the children's raw scores and norms for their age group. These age-corrected scores are the ones used for the present study. Past studies examining the influence of multilingualism on memory generalization have found that bilingual differences are not dependent on exposure to specific language pairs (Brito and Barr 2012, 2014; Brito et al. 2015); therefore, the type of language was not controlled for.

#### 2.2.2. Bilingualism Measure

The Language Exposure Assessment Tool (LEAT; DeAnda et al. 2016) was used to assess language exposure. Parents were interviewed to find out who spoke to the child (i.e., mother, father, sibling, nanny, daycare, etc.), in which language, and for how many hours per day, each day of the week. The percentage of time exposed to each language was calculated for each child from the interview. This measure provided the information needed for the latent profile analysis: language percent exposure (L1, L2, and L3), whether caregivers were native speakers of the child's primary language, and their average proficiency score (0 = not proficient at all to 4 = Very proficient).

##### Binary Language Group Classification (Bilingual = L2 $\geq$ 20%)

A child was classified as bilingual if they were exposed to a second language for 20% or more of their time from birth to the test day (i.e., cumulative bilinguals). Otherwise, the child was classified as monolingual.

##### Group Classification by Latent Profile Analysis

Children's language exposure information gathered using the LEAT was used to identify distinct subgroups (i.e., profiles) of individuals (see the design section for further details).

##### Second Language Exposure (L2%)

Analyses were conducted using the second language exposure estimate calculated in the LEAT.

#### 2.2.3. Stimuli

Small wooden robots (12-cm tall; Tobar) were used (See Figure 1). Each robot can be manipulated to move its limbs and head into different positions (video demonstration can be found in OSF, https://osf.io/q6wfn/ accessed on 12 October 2022). The limbs and head of the robot are connected to the body with an elastic cord, and the body contains spaces next to each limb and head, allowing for movement into the joints. The limbs and head of the robot can move into a total of 16 different positions. At each session, children saw three robots out of six. The selected robots were functionally identical but varied in color, head shape, foot shape, and body markings.

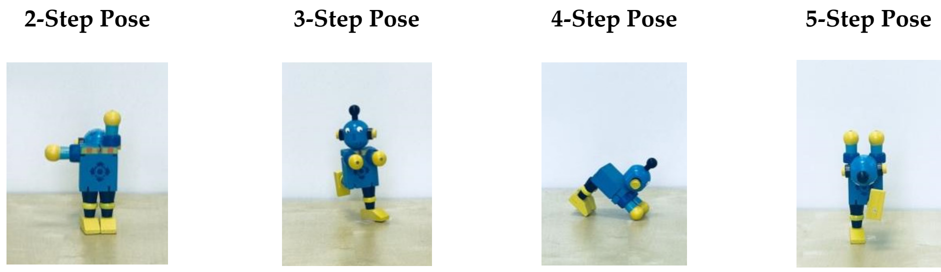


**Figure 1.** Final robot Poses for 2-, 3-, 4- and 5-Step Sequences.

## *2.3. Design and Procedure*

Children participated in an open area in their own homes facing the experimenter. Both the child and the experimenter were seated on the floor. The visit was performed either in Spanish or English by a native speaker based on children's current LEAT scores and parent–child preferences. After a 60-s baseline session and training, there were 2 phases presented in the same order for each sequence: demonstration and test, followed by a manipulation check. On the day of the visit, the child saw three perceptually different robots: robot A during baseline, robot B during training and test, and robot C, which the experimenter used to demonstrate the target actions. Robots were counterbalanced across conditions.

### 2.3.1. Baseline Phase

The experimenter placed robot A on the floor in front of the child and said, "It's your turn!" Children were then given 60 s from the time they first touched the robot to interact with it. We used 60 s for the baseline period because this corresponds to the average length a child spent imitating the most challenging pose (5-step pose).

### 2.3.2. Training Phase

Right after the baseline phase, children received a one-time training session. The experimenter demonstrated how to move robot B's left arm forward twice and invited the participant to imitate the movement twice by stating, "Now it's your turn! Can you show me what I showed you?" The experimenter also demonstrated how to move robot B's right leg to the side twice and asked the participant to imitate the movement twice. This was done to ensure the participant could perform the movements. One child did not participate in the training phase but did not differ from the children who did receive the training on test performance.

### 2.3.3. Demonstration Phase

Participants were seated approximately 50 cm from the experimenter. The experimenter used robot C for each different pose demonstration. The experimenter performed the demonstration three times, ensuring to capture the child's attention throughout. The experimenter reset the robot between demonstrations outside the child's visual field. During the demonstration phase, the experimenter made nonspecific, scripted comments to keep the child engaged in the task (e.g., "Look at this!", "Isn't that fun?"). Four sequences were demonstrated (see Table 1, Figure 1).

**Table 1.** The 2- to 5-step sequences used.

| Pose | Sequence of Target Actions | Test Length (Min-Max) | Max Score |
|---|---|---|---|
| 2-step | 1. Head back<br>2. Left-arm up | 20–30 s | 2 target actions + 1 pair = 3 |
| 3-step | 1. Right arm forward<br>2. Left-arm forward<br>3. Right-leg back | 30–40 s | 3 target actions + 2 pairs = 5 |
| 4-step | 1. Left-leg forward<br>2. Head forward<br>3. Left-arm up<br>4. Right-arm up | 40–50 s | 4 target actions + 3 pairs = 7 |
| 5-step | 1. Head back<br>2. Right-arm up<br>3. Left-arm up<br>4. Left-leg forward<br>5. Right-leg forward | 50–60 s | 5 target actions + 4 pairs = 9 |

#### 2.3.4. Test Phase

Children were tested using robot B immediately after each pose demonstration. The experimenter placed robot B on the floor in front of the child and asked, "Can you show me what I showed you?" For each test phase, the child was given a robot to imitate the different poses. That is, they had to update the sequence on robot B but match the actions to a demonstration on robot C. Children were tested on 2-, 3-, 4-, and 5-step sequences, and test time varied as a function of the number of steps (see Table 1 and Rusnak et al. 2022). If the child completed the pose before the allotted time, the experimenter reset the robot outside the child's visual field and asked the child to recreate the pose. Children did not receive feedback on the accuracy of their imitation.

#### 2.3.5. Manipulation Check

At the end of the last sequence's test, if the child did not complete movements for the sequence, the experimenter demonstrated each pose one time and asked the child to imitate the target actions, guiding the child through the pose. This ensured that the child possessed the motoric ability to complete each pose; all children did.

### *2.4. Coding and Dependent Variables*

All sessions were video recorded for later coding. Coders converted each video using Movavi Suite (https://www.movavi.com/ accessed on 12 October 2022) and ffmpegX (https://www.ffmpegx.com/ accessed on 12 October 2022) software to a format compatible with the Datavyu (http://datavyu.org/ accessed on 12 October 2022) software. Coding in Datavyu consisted of timestamping each movement, indicating the piece, orientation, and any comments associated with the movement or the entire task. All statistical analyses were conducted in R version 4.1.1 (8 October 2021).

#### 2.4.1. Total Correct Movements Score

The total number of correct pieces in the correct orientation at the end of the test phase.

#### 2.4.2. Pair Score

This score refers to whether children followed a correct sequence of two steps, with the maximum possible correct pair score depending on the pose. The possible range of scores is 0 to 1 pair for a 2-step sequence, 0 to 2 pairs for a 3-step sequence, 0 to 3 pairs for a 4-step sequence, and 0 to 4 pairs for a 5-step sequence.

2.4.3. Composite Percentage Score

We combined the total correct movements plus the number of pairs that children achieved to create a composite score for each sequence. We divided the composite score by the total possible maximum composite score at each sequence to produce a composite score and multiplied it by 100 to ease visualization and interpretation.

$$Composite\ Percentage\ Score = \frac{Total\ Correct\ Movements + Total\ Pair\ Achieved}{Total\ Posible\ Movements + Total\ Possible\ Pairs} * 100$$

2.4.4. Baseline Composite Percentage Score

Like the test trials, the 60-s baseline session was coded for the pieces moved and the orientation of all the movements children produced. Then, we determined whether any of the movements and pairs of movements that children spontaneously produced corresponded to any of the sequences demonstrated during test trials. This was quite a conservative approach, given that this estimate included all the sequences in each variant that were demonstrated and any spontaneous movement could occur in more than one sequence. That is, the same spontaneous production of a movement in a particular orientation could contribute to the baseline calculation for more than one sequence. For example, a child that moves the robot's head to the back during the baseline phase would count as a correct movement for two poses (2-step and 4-step poses).

### *2.5. Latent Profile Analysis (LPA) Plan*

Two Latent Profile Analyses were performed to find out the best number of latent profiles based on children's language exposure information with the sample from each approach: L1% exposure, L2% exposure, L3% exposure, at least one parent native in the child's L1 (Yes = 1; No = 0), at least one parent native in child's L2 (Yes = 1; No = 0), average parental fluency in child's L1 (0–4), Code Switching (0–30). For LPA, we looked at the model fit of identifying two to six latent profiles. Based on the fit indices AIC, AWE, BIC, CLC, and KIC (Akogul and Erisoglu 2017), an analytic hierarchy process suggested the best solutions for the samples were Model 1 with three classes. The three classes were labeled as Low, Medium, and High multilingual exposure (See Figure 2 of LPA with the sample from Approach 1).

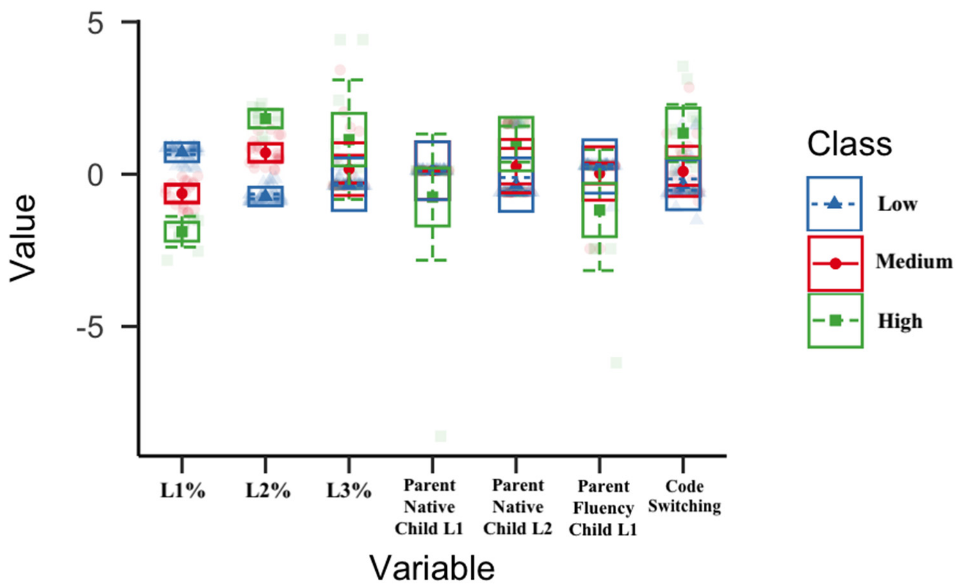


**Figure 2.** Latent Profile Plot for Analysis Approach 1. Variables were scaled for visualization purposes.

### 2.6. Analysis Plan

Two analysis approaches were formulated to assess the generalization skills of children growing up exposed to one or more languages.

1. For approach one, a summary of performance across 2- to 5-step sequences was calculated. This included the number of target actions and pairs the child achieved, divided by the possible maximum score. To be included in the three linear regression analyses, children must have completed the baseline phase and all four poses. Of the 101 children, 76 had complete data for all poses and thus a score for each sequence length. There was no differential dropout rate as a function of bilingual status. The full models were:

- *Model 1: Binary Language Group Classification (Bilingual = L2 $\geq$ 20%)*

total_composite_percentage_score ~ lang_group (Monolingual vs. Bilingual) + L3%_exposure + child_sex + age_centered + edu_centered + income_centered + pvt_perc

- *Model 2: Group Classification by Latent Profile Analysis*

total_composite_percentage_score ~multilingual_exposure (Low vs. Medium vs. High) + child_sex + age_centered + edu_centered + income_centered + pvt_perc

- *Model 3: Second Language Exposure (L2%)*

total_composite_percentage_score ~ L2%_exposure + L3%_exposure + child_sex + age_centered + edu_centered + income_centered + pvt_perc

2. For approach two, we conducted three multi-level mixed-effects models to account for the clustering of observations within subjects. Children were included in the dataset if they had completed at least one valid trial. We used a random intercept to allow children's composite scores across conditions to vary. The full models were:

- *Model 1: Binary Language Group Classification (Bilingual = L2 $\geq$ 20%)*

total_composite_percentage_score ~ lang_group (Monolingual vs. Bilingual) + L3%_exposure + child_sex + age_centered + edu_centered + income_centered + seq_length + pvt_perc + (1 | subj)

- *Model 2: Group Classification by Latent Profile Analysis*

total_composite_percentage_score ~ multilingual_exposure (Low vs. Medium vs. High) + child_sex + age_centered + edu_centered + income_centered + seq_length + pvt_perc + (1 | subj)

- *Model 3: Second Language Exposure (L2%)*

total_composite_percentage_score ~L2%_exposure + L3%_exposure + child_sex + age_centered + edu_centered + income_centered + seq_length + pvt_perc + (1 | subj)

For each analysis, a full model and a reduced model (without the education and income covariates) were compared using Akaike values (AIC) for better fit.

## 3. Results

### 3.1. Inter-Coder Reliability

A primary coder for each child and each test was designated. A secondary coder coded the video for reliability purposes using the timestamps for each movement coded by the primary coder. For each timestamp, the reliability coder coded the piece, and the orientation moved. The videos of 36% of the participants were coded for reliability purposes. Inter-coder agreement on the piece and orientation ($k_{piece}$ = 0.89, $k_{orientation}$ = 0.93) were in the acceptable range above 0.70 (Landis and Koch 1977).

### 3.2. Vocabulary Analysis

There was no significant difference on PVT Percentile scores across all language groups or significant correlations with Composite Score at either Baseline or Test for the Approach 1, respectively, ($r(61) = 0.19$, $p = 0.14$; $r(61) = 0.19$, $p = 0.13$). Additionally, there was no significant correlation between L2% exposure for either sample (see Table 2 for more details and descriptive analyses).

**Table 2.** PVT Percentile Score Analyses for All Analytical Approaches.

| | | Variables | *M* | *SD* | Min | Max | Statistic |
|---|---|---|---|---|---|---|---|
| Sample from Approach 1 | Binary | Monolingual<br>Bilingual | 45.21<br>37.54 | 23.83<br>23.48 | 1<br>2 | 91<br>78 | $t(29.91) = 0.65$,<br>$p = 0.52$ |
| | Profiles | Low<br>Medium<br>High | 46.85<br>39.20<br>39.89 | 22.88<br>24.91<br>21.97 | 6<br>2<br>6 | 91<br>78<br>78 | $F(2, 60) = 0.752$,<br>$p = 0.476$ |
| | Continuous | PVT Percentile<br>L2% | 44.03<br>12.61 | 23.16<br>14.19 | 2<br>0 | 91<br>45.50 | $r(61) = -0.107$,<br>$p = 0.401$ |
| Sample from Approach 2 | Binary | Monolingual<br>Bilingual | 45.21<br>37.54 | 23.83<br>23.48 | 1<br>2 | 91<br>78 | $t(44.19) = 1.33$,<br>$p = 0.19$ |
| | Profiles | Low<br>Medium<br>High | 46.06<br>41.90<br>33.08 | 22.85<br>26.54<br>21.92 | 6<br>1<br>2 | 91<br>78<br>78 | $F(2, 77) = 1.556$,<br>$p = 0.217$ |
| | Continuous | PVT Percentile<br>L2% | 42.91<br>13.34 | 23.84<br>14.66 | 1<br>0 | 91<br>45.5 | $r(78) = -0.156$,<br>$p = 0.168$ |

### *3.3. Imitation*

To ensure that preschoolers are imitating newly learned behavior, it is important that their performance is compared to a condition without any demonstrations (baseline condition). We expected that if children were indeed learning a new behavior, they would present a low rate of spontaneous production of the target actions. For children with complete data, the total composite percentage score for the Test condition ranged from 0 to 92% (M = 43.08, SD = 24.07) and 0 to 38 (M = 1.41, SD = 5.56) for the Baseline condition.A t-test confirmed our hypothesis, as children had significant lower composite imitation scores before any demonstrations were performed ($t(75) = -14.91$, $p < 0.00001$).

This finding was also replicated when their performance was analyzed at a trial-by-trial level using the sample from approach 2, which included children with at least one complete trial. Despite the level of steps required for the pose, children performed significantly higher during the test trials than in baseline trials (see Tables S1–S3 of results in the Supplementary Materials). Demonstrating that children were not as likely to come up with the target actions of their own volition without the experimenter demonstration showing that this is a feasible imitation task for this target age group. For the following analyses, we focused on children's performance during the test phase to investigate the relationship between the type of language exposure (using three measures of multilingual exposure) and generalization skills using the Total Composite Score (Approach 1) and the Trial-by-Trial Composite Scores (Approach 2).

### *3.4. Analysis Approach 1: Linear Regression Analysis*

For children with complete data ($N = 76$), the total composite percentage score for the Test condition ranged from 0 to 92% ($M = 43.08$, $SD = 24.07$) and was normally distributed.

#### 3.4.1. Binary Language Group Classification (Bilingual = L2 $\geq$ 20%)

The best fitting model was:

total_composite_percentage_score ~ lang_group (Monolingual vs. Bilingual) + l3%_exposure + age_centered + pvt_percentile

Table 3 shows the coefficient estimates from the best fitting model by bilingualism proxy, and Figure 3 displays them. There was no significant difference between Monolinguals ($N$ = 54, $M$ = 49.14, $SD$ = 22.78) and Bilinguals ($N$ = 22, $M$ = 40.61, $SD$ = 24.34).

**Table 3.** Linear Regression Models for Analysis Approach 1-Monolinguals vs. Bilinguals.

| | **Full Model Estimate (*SE*)** | **Final Model Estimate (*SE*)** |
|---|---|---|
| (Intercept) | 33.366 ** | 33.400 *** |
| | (9.773) | (8.599) |
| Language Group (Monolingual vs. Bilingual) | −2.699 | −4.035 |
| | (8.046) | (7.196) |
| L3% exposure | 0.955 | 0.847 |
| | (0.810) | (0.784) |
| Child's Age in Months (Centered) | 1.921 | 1.966 |
| | (1.320) | (1.277) |
| PVT Percentile (Vocabulary) | 0.197 | 0.204 |
| | (0.138) | (0.135) |
| Child's Sex (Male = 0; Female = 1) | −1.754 | |
| | (7.003) | |
| Avg. Parental Education (Centered) | −2.320 | |
| | (2.776) | |
| Income (Centered) | $-6.384 \times 10^{-5}$ | |
| | $(1.192 \times 10^{-4})$ | |
| Num. Obs. | 63 | 63 |
| *R*2 | 0.130 | 0.115 |
| *R*2 Adj. | 0.019 | 0.054 |
| AIC | 591.1 | 586.2 |
| BIC | 610.4 | 599.0 |
| Log. Lik. | −286.553 | −287.087 |
| *F* | 1.169 | 1.878 |
| RMSE | 24.47 | 24.03 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. ** $p < 0.01$, *** $p < 0.001$.

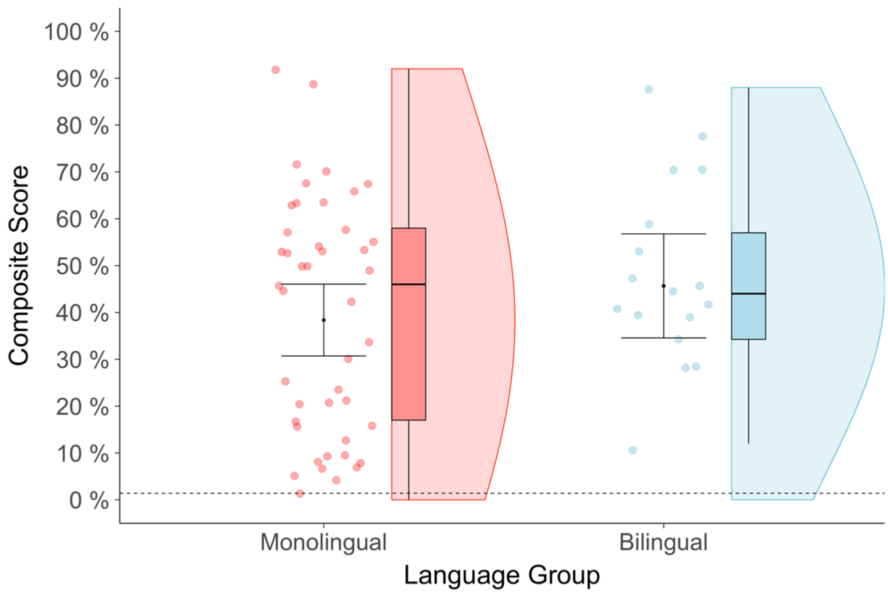


**Figure 3.** Distribution of Total Composite Scores by Condition and Language Group. The dashed line represents the average mean composite score in the Baseline condition ($M$ = 1.41, $SD$ = 5.56).

### 3.4.2. Model 2. Group Classification by Latent Profile Analysis

The best-fitting model was:

total_composite_percentage_score ~ multilingual_exposure (Low vs. Medium/Low vs. High) + L3%_exposure + age_centered + pvt_percentile

A total of 45 children were clustered in the Low multilingual exposure group, 21 in the Medium multilingual exposure group, and ten were classified in the High multilingual exposure group. Table 4 shows the coefficient estimates from the best fitting model with the identified profiles as the main predictor, and Figure S1 displays them. There was no significant difference among the multilingual exposure profiles (Low vs. Medium/ Low vs. High) or the added covariates (age, vocabulary percentile scores, and L3% exposure).

**Table 4.** Linear Regression Models for Analysis Approach 1-Latent Profile Analysis.

| | **Full Model Estimate (*SE*)** | **Final Model Estimate (*SE*)** |
|---|---|---|
| (Intercept) | 27.305 ** | 26.111 ** |
| | (8.246) | (7.570) |
| Multilingual Exposure Profiles | 11.807 | 11.402 |
| (Low vs. Medium) | (8.014) | (7.744) |
| Multilingual Exposure Profiles | 3.847 | 4.776 |
| (Low vs. High) | (11.210) | (10.534) |
| L3% Exposure | 0.829 | 0.719 |
| | (0.897) | (0.871) |
| Child's Age in Months (Centered) | 1.511 | 1.611 |
| | (1.333) | (1.293) |
| PVT Percentile (Vocabulary) | 0.226 | 0.230 + |
| | (0.138) | (0.135) |
| Child's Sex (Male = 0; Female = 1) | −2.607 | |
| | (6.914) | |
| Avg. Parental Education (Centered) | −2.730 | |
| | (2.764) | |
| Income (Centered) | $-6.344 \times 10^{-5}$ | |
| | $(1.127 \times 10^{-4})$ | |
| Num. Obs. | 63 | 63 |
| *R*2 | 0.162 | 0.142 |
| *R*2 Adj. | 0.038 | 0.067 |
| AIC | 590.7 | 586.2 |
| BIC | 612.1 | 601.2 |
| Log. Lik. | −285.353 | −286.080 |
| F | 1.305 | 1.894 |
| RMSE | 24.23 | 23.86 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. + $p < 0.1$, ** $p < 0.01$.

### 3.4.3. Model 3. Second Language Exposure (L2%)

The best-fitting model was:

total_composite_percentage_score ~ L2%_exposure + L3%_exposure + age_centered + pvt_percentile

Table 5 shows the coefficient estimates from the best fitting model with the child's L2 exposure as our continuous main predictor and Figure S2 displays it. Neither L2 exposure nor covariates were significant predictors of children's generalization performance.

**Table 5.** Regression Models for Analysis Approach 1-L2% Exposure.

| | **Full Model Estimate (*SE*)** | **Final Model Estimate (*SE*)** |
|---|---|---|
| (Intercept) | 29.844 *** | 28.656 *** |
| | (8.292) | (7.503) |
| L2% Exposure | 0.131 | 0.155 |
| | (0.253) | (0.237) |
| L3% Exposure | 0.878 | 0.782 |
| | (0.828) | (0.805) |
| Child's Age in Months (Centered) | 1.891 | 1.960 |
| | (1.315) | (1.273) |
| PVT Percentile (Vocabulary) | 0.200 | 0.206 |
| | (0.138) | (0.135) |
| Child's Sex (Male = 0; Female = 1) | −1.774 | |
| | (6.932) | |
| Avg. Parental Education(Centered) | −2.383 | |
| | (2.776) | |
| Income (Centered) | −0.0000626 | |
| | (0.0001139) | |
| Num. Obs. | 63 | 63 |
| *R*2 | 0.132 | 0.116 |
| *R*2 Adj. | 0.022 | 0.055 |
| AIC | 590.9 | 586.0 |
| BIC | 610.2 | 598.9 |
| Log. Lik. | −286.464 | −287.025 |
| *F* | 1.195 | 1.910 |
| RMSE | 24.43 | 24.01 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. *** $p < 0.001$.

### *3.5. Analysis Approach 2: Multi-Level Modeling for Trial-by-Trial Performance*

For children in the final dataset who contributed with at least one valid test trial ($N = 100$), the total composite percentage score for the Test condition ranged from 0 to 100% ($M = 43.59$, $SD = 36.37$).

#### 3.5.1. Model 1. Binary Language Group Classification (Bilingual = L2 ≥ 20%)

The best-fitting model was:

total_composite_percentage_score ~ bilingual_group (Monolingual vs. Bilingual) + L3%_exposure + child_sex + age_centered + seq_length + (1 | subj)

Table 6 shows the coefficient estimates from the best fitting model by bilingualism proxy, and Figure 4 displays them. There was no significant difference between Monolinguals and Bilinguals. Additionally, we could predict a 3.8% reduction in their test composite score for any additional step added to the sequence.

**Table 6.** Mixed Effects Models for Analysis Approach 2-Binary Language Classification.

| | **Full & Final Model Estimate (SE)** |
|---|---|
| (Intercept) | 44.797 *** <br> (8.117) |
| Language Group (Monolingual vs. Bilingual) | 0.839 <br> (6.796) |
| L3% Exposure | 0.565 <br> (0.688) |
| Child's Sex (Male = 0; Female = 1) | −7.193 <br> (5.819) |
| Length of Sequence | −3.776 * <br> (1.645) |
| Avg. Parental Education (Centered) | −2.410 <br> (2.241) |
| Income (Centered) | $-2.206 \times 10^{-5}$ <br> $(9.998 \times 10^{-5})$ |
| Child's Age in Months (Centered) | 1.996 <br> (1.220) |
| PVT Percentile (Vocabulary) | 0.174 <br> (0.120) |
| *SD* (Intercept) | 17.744 |
| *SD* (Observations) | 30.678 |
| Num. Obs. | 290 |
| *R*2 Marg. | 0.057 |
| *R*2 Cond. | 0.293 |
| AIC | 2893.4 |
| BIC | 2933.7 |
| RMSE | 28.31 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. * $p < 0.05$, *** $p < 0.001$.

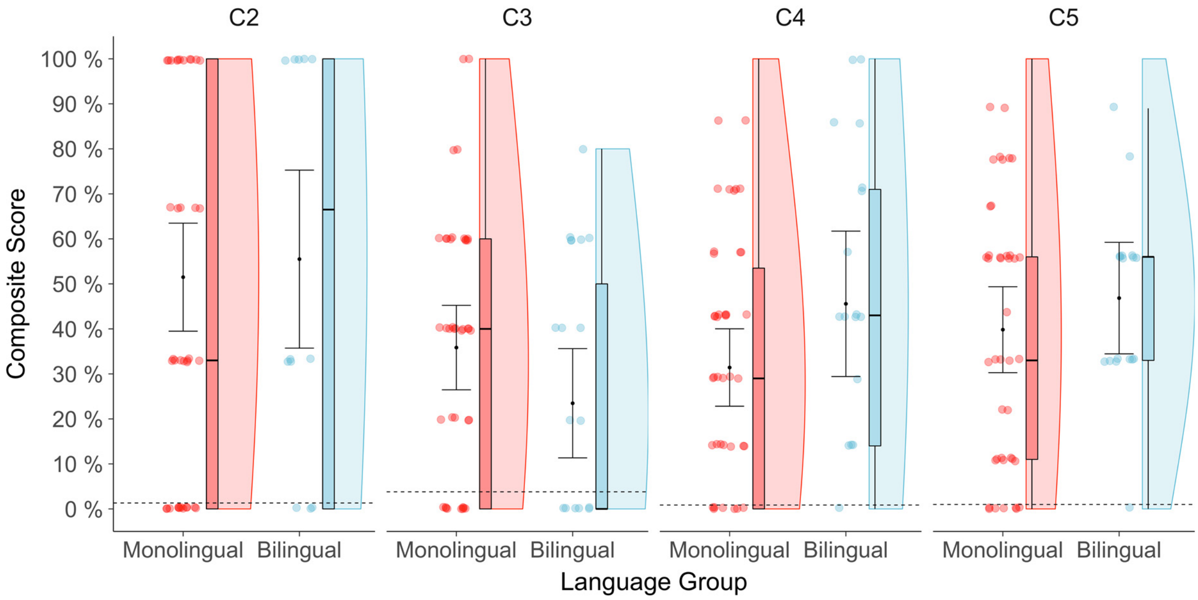


**Figure 4.** Distribution of Composite Scores by Condition and Pose. The average mean composite score in the Test condition per sequence was: $M_{2\text{-}step}$ = 57.26%; $M_{3\text{-}step}$ = 33.75%; $M_{4\text{-}step}$ = 38.45%; $M_{5\text{-}step}$ = 44.38%) The horizontal dashed lines represent the average mean composite score in the Baseline condition ($M_{2\text{-}step}$ = 1.32%; $M_{3\text{-}step}$ = 3.80%; $M_{4\text{-}step}$ = 0.86%; $M_{5\text{-}step}$ = 0.99%).

#### 3.5.2. Model 2. Group Classification by Latent Profile Analysis

The best-fitting model was:

total_composite_percentage_score ~ multilingual_exposure (Low vs. Medium/ Low vs. High) + L3%_exposure + pvt_percentile + age_centered + seq_length + (1 | subj)

A total of 58 children were clustered in the Low multilingual exposure group, 27 in the Medium multilingual exposure group, and 16 were classified in the High multilingual exposure group. Table 7 shows the coefficient estimates from the best fitting model with the identified profiles as the main predictor, and Figure S3 displays them. There was no significant difference among the multilingual exposure profiles (Low vs. Medium/Low vs. High). Additionally, for every added step to the pose in the Test phase, we could predict a 3.4% reduction in their composite score.

**Table 7.** Mixed Effects Models for Analysis Approach 3-Latent Profile Analysis.

| | **Full Model Estimate (*SE*)** | **Final Model Estimate (*SE*)** |
|---|---|---|
| (Intercept) | 44.112 *** | 41.248 *** |
| | (8.319) | (7.889) |
| Multilingual Exposure Profiles | 3.061 | 2.218 |
| (Low vs. Medium) | (6.724) | (6.776) |
| Multilingual Exposure Profiles | −0.240 | −1.851 |
| (Low vs. High) | (9.566) | (9.240) |
| L3% Exposure | 0.598 | 0.623 |
| | (0.756) | (0.765) |
| Child's Age in Months (Centered) | 1.932 | 1.904 |
| | (1.229) | (1.229) |
| PVT Percentile (Vocabulary) | 0.174 | 0.171 |
| | (0.120) | (0.120) |
| Length of Sequence | −3.765 * | −3.715 * |
| | (1.646) | (1.647) |
| Child's Sex (Male = 0; Female = 1) | −7.039 | |
| | (5.801) | |
| Avg. Parental Education (Centered) | −2.442 | |
| | (2.244) | |
| Income (Centered) | $-2.910 \times 10^{-5}$ | |
| | $(9.859 \times 10^{-5})$ | |
| *SD* (Intercept) | 17.654 | 18.081 |
| *SD* (Observations) | 30.691 | 30.704 |
| Num. Obs. | 290 | 290 |
| *R*2 Marg. | 0.058 | 0.046 |
| *R*2 Cond. | 0.292 | 0.292 |
| AIC | 2895.1 | 2891.4 |
| BIC | 2939.2 | 2924.4 |
| RMSE | 28.34 | 28.30 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. * $p < 0.05$, *** $p < 0.001$.

#### 3.5.3. Model 3. Second Language Exposure (L2%)

The best-fitting model was:

total_composite_percentage_score ~ L2%_exposure + L3%_exposure + age_centered + seq_length + pvt_percentile + (1 | subj)

Table 8 shows the coefficient estimates from the best fitting model with child's L2 exposure as our continuous main predictor, and Figure S4 displays it. L2 exposure was not a significant predictor of children's generalization performance. Additionally, we could predict a 3.72% reduction in their composite scores for every step added to the sequence during the Test phase.

**Table 8.** Regression Models for Analysis Approach 3 -L2% Exposure.

| | Full Model Estimate (*SE*) | Final Model Estimate (*SE*) |
|---|---|---|
| (Intercept) | 44.821 *** | 41.856 *** |
| | (8.413) | (7.949) |
| L2% Exposure | 0.017 | −0.014 |
| | (0.216) | (0.211) |
| L3% Exposure | 0.573 | 0.562 |
| | (0.694) | (0.702) |
| Length of Sequence | −3.775 * | −3.721* |
| | (1.645) | (1.646) |
| Child's Age in Months (Centered) | 2.006 | 1.959 |
| | (1.221) | (1.219) |
| PVT Percentile (Vocabulary) | 0.173 | 0.170 |
| | (0.120) | (0.120) |
| Child's Sex (Male = 0; Female = 1) | −7.123 | |
| | (5.774) | |
| Avg. Parental Education (Centered) | −2.405 | |
| | (2.247) | |
| Income (Centered) | $-2.405 \times 10^{-5}$ | |
| | $(9.794 \times 10^{-5})$ | |
| *SD* (Intercept) | 17.749 | 18.155 |
| *SD* (Observations) | 30.678 | 30.694 |
| Num. Obs. | 290 | 290 |
| *R*2 Marg. | 0.057 | 0.045 |
| *R*2 Cond. | 0.293 | 0.292 |
| AIC | 2893.4 | 2889.6 |
| BIC | 2933.7 | 2919.0 |
| RMSE | 28.31 | 28.28 |

*Note.* The income variable is based on the median household income zip code for the family's postcode at the time of participation. * $p < 0.05$, *** $p < 0.001$.

### *3.6. Exploratory Analysis*

Following Rusnak et al. (2022) 's protocol, we analyzed children's cognitive load performance during the test trials. 2- and the 3-step-sequences were classified as Low Load trials, and 4- and 5-step-sequences were classified as High Load trials. Following the previous analyses, we explored the relationship between bilingualism and generalization performance using a binary language classification, a latent profile approach, and a continuum measure of second language exposure (See Supplementary Tables S4–S6 and Figure 5 for visualization). Overall, children scored significantly worse in the higher load trials compared to the low load trials, but there was no significant difference between language groups or L2 percent exposure.

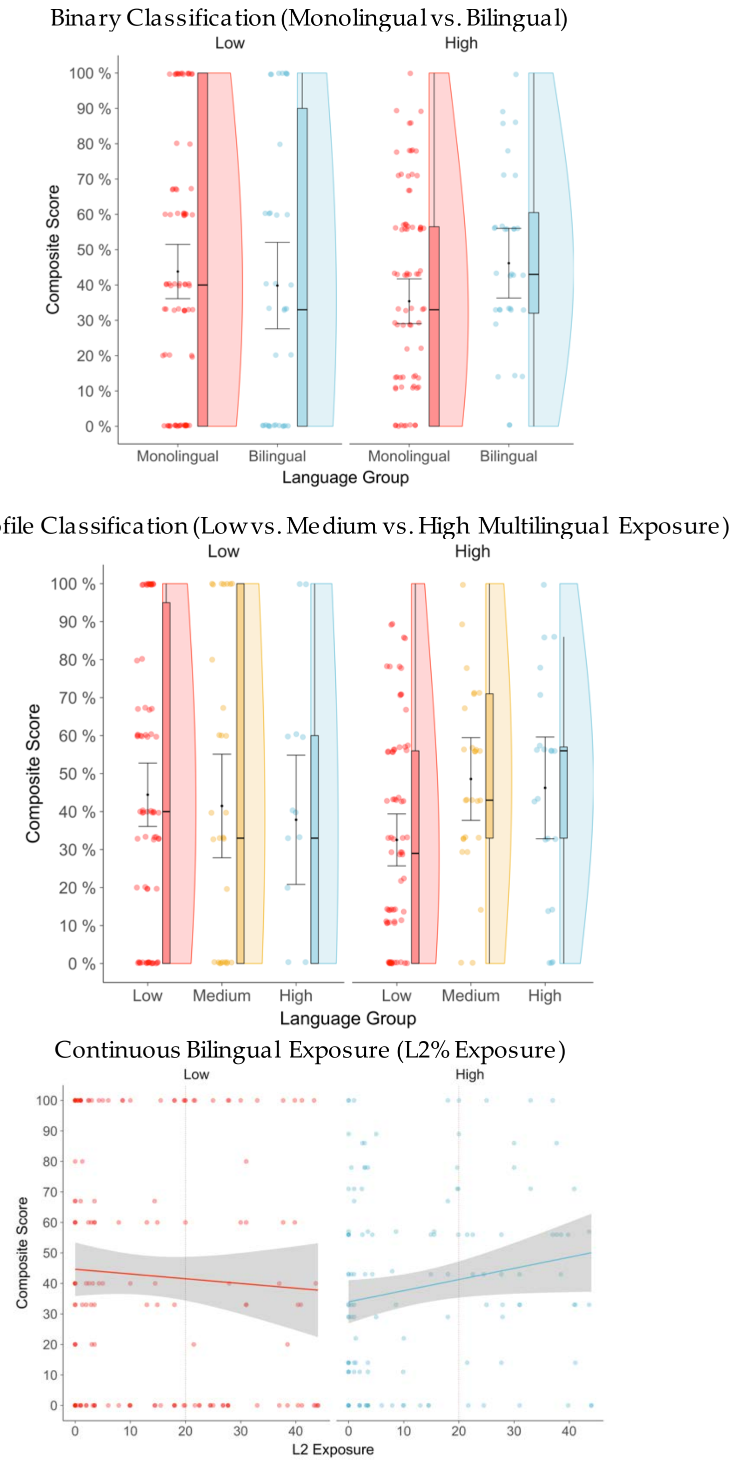


**Figure 5.** Exploratory Analysis Measuring Load and Language Exposure. Three approaches to measuring language exposure were used: a binary classification of Monolinguals and Bilinguals (Top Panel), profiles using a Latent Profile Analysis approach (Middle Panel), and bilingualism as a continuum using the L2 percent exposure as a proxy (Bottom Panel).

## 4. Discussion

We used the object sequencing imitation task (Rusnak et al. 2022) to test memory flexibility and working memory in monolingual and bilingual 3-year-olds. We demonstrated sequences on one wooden robot figurine and tested children on a functionally similar but perceptually different robot. We hypothesized that all children would perform above baseline and that bilingual children would perform significantly better than monolingual children. Both monolingual and bilingual preschoolers performed significantly above baseline but did not differ on the total composite percentage score, which was a summary of performance across 2- to 5-step sequences and included the number of target actions and pairs children achieved, divided by the possible maximum score. The baseline score was low, nearly zero, which is a hallmark of a robust imitation task.

Data were also analyzed using a multi-level model to include all participants who completed two or more trials and increase the analytic power. Consistent with the across-trials composite score model, we found that preschoolers performed significantly better on the test than the baseline, that performance on the test decreased as a function of the increasing sequence length, that performance increased as a function of age, and that bilingual 3-year-olds did not differ from monolinguals on the task.

One goal of the present study was to examine whether it was feasible to examine different patterns of language exposure using multiple methods, a traditional binary categorization, a continuous %L2 measure, and a latent profile analysis. We achieved this goal. We found that for this study, the findings converged across each approach. Future studies should also examine whether the results pattern converges or differs across different approaches to determine better and more consistent practice in bilingual definitions during early childhood (Rocha-Hidalgo and Barr 2022). We could have taken other approaches to bilingual categorization, such as looking more closely at the extremes of the sample but based on the pattern of the results, we do not think that our findings would differ.

We found that there were no differences across our three analytic approaches on the OSI task between children growing up in monolingual and bilingual households. There are a number of possible explanations for these findings. We describe the task as both a visual-spatial working memory task (Rusnak et al. 2022 and as a test of memory flexibility. That is, the task includes elements of executive functioning and representational flexibility. In terms of executive functioning, the lack of differences in performance as a function of language status is consistent with recent meta-analyses. For example, Lowe and colleagues (Lowe et al. 2021) reported no significant differences in executive functioning in 3- to 17-year-olds. Their study excluded infants, toddlers, and adults. However, the classification of bilinguals differed across the ages making it more difficult to interpret the findings (Rocha-Hidalgo and Barr 2022). Gunnerud et al. (2020) conducted a systematic review of children 18 years and under. They reported marginal differences between monolinguals and bilinguals, particularly in code-switching and inhibition, but no differences with monitoring. The authors reported that these findings might have been affected by publication bias. Finally, Beaudin and Poulin-Dubois (2022) reported in a series of studies of toddlers and preschoolers that any differences between monolinguals and bilinguals were limited to differences in inhibitory control. Although the OSI task requires updating, it does not explicitly measure inhibitory control, and therefore our findings are consistent with Beaudin and Poulin-Dubois' conclusions.

We suggest an alternate interpretation of our data. Studies examining differences in executive functioning have not considered precursors to executive functioning in representational processing, such as differences in memory flexibility. Although there were no significant bilingual differences in the present study, there was a significant degree of individual variability in performance. Changes in memory flexibility may be a precursor to later differences in cognitive flexibility—the ability to adjust to changes in task demands and switch between different rules and goals (Mahy and Munakata 2015)—that have been reported between monolingual and bilingual children. Bilingual language status is associated with earlier trajectories in memory flexibility during infancy (e.g., Brito and Barr 2012) and cognitive flexibility in preschool-aged children (Adi-Japha et al. 2010; Bialystok and Martin 2004; Bialystok and Senman 2004; Carlson and Meltzoff 2008) and throughout the lifespan (Bialystok et al. 2006; Costa et al. 2008). We hypothesize that there might be different time points when cognitive processing trajectories diverge between monolinguals and bilinguals. We argue that the current study suggests that the trajectory for memory flexibility may be converging with monolingual and bilingual differences disappearing by three years of age. That is, on average, monolinguals are improving in memory flexibility. Both groups are about 50% on the composite; this shows this task was challenging. However, we also compared high and low load performance, and we did not see a difference there either. That is, based on our range of scores, the lack of differences between monolinguals and bilinguals should not be attributed to floor or ceiling effects. At three years of age, it is possible that differences in inhibitory control may be the most apparent cognitive differences, as argued by Beaudin and Poulin-Dubois (2022).

Our contention requires empirical evaluation via longitudinal analysis, which our group is currently evaluating. There may be age-related changes in how systems differ between monolinguals and bilinguals. The collected data for the present study are part of an ongoing longitudinal study examining language exposure, memory flexibility, and cognitive flexibility in 1- to 5-year-olds. The individual differences exhibited in this task make it particularly valuable to track changes in memory flexibility as language patterns shift over time (Rocha-Hidalgo et al. forthcoming). For example, children may hear more of a second language before entering childcare or preschool. After entering preschool, their exposure to their primary and secondary languages may shift. The OSI task can test memory flexibility across the preschool years as it allows us to manipulate memory load via the number of items to remember, and data collection with 5-year-olds is ongoing. Performance on the OSI task will be compared to performance on imitation tasks at younger ages to examine whether earlier imitation performance measured by an interference task at 18 months and generalization tasks at 24 months predict later performance and whether performance at three years is related to performance at five years of age.

There were, however, limitations to the present study. The OSI task differs from prior memory generalization tasks used to measure memory flexibility in infants and toddlers. It was by design that we created a task that required multiple updates and hence flexibility across poses. Still, it is possible that children also rapidly overcame any perceptual differences between the robots. Additional research using additional memory flexibility tasks will be needed to test whether there are ongoing perceptual processing or selective attention differences between monolinguals and bilingual preschoolers. This could be achieved by taking advantage of the flexibility of the OSI task by increasing the perceptual differences between the different objects.

Although we recruited widely, the distribution of our sample was skewed toward monolinguals, and there were few balanced bilinguals, which may have reduced our ability to fully exploit the utility of latent profile analysis. These data could also be subjected to Bayesian analysis to determine the probability of the null findings but based on the pattern of results; we would expect that these null findings are robust. There was significant variability in performance which will be helpful in future longitudinal data analyses. However, one potential source of unexplained variance may be because some data were collected during the COVID pandemic and health guidelines meant that experimenters were masked during the session. It is too early to know how these differences may be related to imitation performance and whether there will be differences in social learning in children born after the pandemic. Statistically, we did not have the power to analyze this difference in our dataset.

## 5. Conclusions

Taken together, this study demonstrates memory flexibility and working memory performance in a group of 3-year-olds who were very well characterized for language exposure. Multiple methods were used to analyze language exposure differences, and the results converged, showing no bilingual differences. There was significant individual variability in this task, and these data will be further examined in longitudinal analyses.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/languages7040268/s1.

**Author Contributions:** Conceptualization, J.R.-H., S.N.R. and R.B.; Data curation, J.R.-H., O.A.B., L.T. and K.C.; Formal analysis, J.R.-H. and S.N.R.; Funding acquisition, R.B.; Investigation, J.R.-H., S.N.R., O.A.B. and K.C.; Methodology, J.R.-H., S.N.R. and R.B.; Project administration, J.R.-H., O.A.B., K.C. and R.B.; Supervision, J.R.-H., O.A.B., K.C. and R.B.; Validation, J.R.-H., S.S., L.T. and H.G.; Visualization, J.R.-H.; Writing—original draft, J.R.-H., S.N.R., O.A.B. and R.B.; Writing—review and editing, J.R.-H., S.N.R., O.A.B., S.S., L.T., H.G., K.C. and R.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the NSF BCS-1551719.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Georgetown University (protocol code CR00002477 and 11/23/2022 Continuing Review approval date).

**Informed Consent Statement:** Informed consent was obtained from a parent of all subjects involved in the study.

**Data Availability Statement:** Data supporting reported results and scripts for the analyses can be found at https://osf.io/q6wfn/ accessed on 12 October 2022.

**Acknowledgments:** We are grateful to all the families who participated in this research and our Research Assistants: Monica Collins, Christine Kim, Subul Malik, and Tuleen Sawaf, for their invaluable help.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

Adi-Japha, Esther, Jennie Berberich-Artzi, and Afaf Libnawi. 2010. Cognitive flexibility in drawings of bilingual children. *Child Development* 81: 1356–66. [CrossRef] [PubMed]

Akogul, Serkan, and Murat Erisoglu. 2017. An approach for determining the number of clusters in a model-based cluster analysis. *Entropy* 19: 452. [CrossRef]

Baddeley, Alan D., and Graham Hitch. 1974. Working memory. In *The Psychology of Learning and Motivation: Advances in Research and Theory*. Edited by Gordon H. Bower. New York: Academic Press, vol. 8, pp. 47–89.

Barac, Raluca, Ellen Bialystok, Dina C. Castro, and Marta Sanchez. 2014. The Cognitive Development of Young Dual Language Learners: A Critical Review. *Early Childhood Research Quarterly* 29: 699–714. [CrossRef] [PubMed]

Barr, Rachel, and Natalie Brito. 2014. From specificity to flexibility: Early developmental changes in memory generalization. In *The Wiley Handbook on the Development of Children's Memory*. Edited by Patricia J. Bauer and Robyn Fivush. Chichester: John Wiley and Sons, pp. 453–79.

Barr, Rachel, and Harlene Hayne. 2000. Age-related changes in imitation: Implications for memory development. In *Progress in Infancy Research*. Edited by Carolyn Rovee-Collier, Lewis P. Lipsitt and Harlene Hayne. Mahwah: Erlbaum, vol. 1, pp. 21–67.

Barr, Rachel, Anne Dowden, and Harlene Hayne. 1996. Developmental changes in deferred imitation by 6-to 24-month-old infants. *Infant Behavior and Development* 19: 159–70. [CrossRef]

Barr, Rachel, Alecia Moser, Sylvia Rusnak, Laura Zimmermann, Kelly Dickerson, Herietta Lee, and Peter Gerhardstein. 2016. The impact of memory load and perceptual cues on puzzle learning by 2-year-olds. *Developmental Psychobiology* 58: 817–28. [CrossRef] [PubMed]

Barr, Rachel, Sylvia N. Rusnak, Natalie H. Brito, and Courtney Nugent. 2019. Actions speak louder than words: Differences in memory flexibility between monolingual and bilingual 18-month-olds. *Developmental Science* 23: e12881. [CrossRef]

Beaudin, Kayla, and Diane Poulin-Dubois. 2022. Testing the Bilingual Cognitive Advantage in Toddlers Using the Early Executive Functions Questionnaire. *Languages* 7: 122. [CrossRef]

Bialystok, Ellen, and Michelle M. Martin. 2004. Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science* 7: 325–39. [CrossRef]

Bialystok, Ellen, and Lili Senman. 2004. Executive processes in appearance–reality tasks: The role of inhibition of attention and symbolic representation. *Child Development* 75: 562–79. [CrossRef]

Bialystok, Ellen, Fergus I. Craik, and Jennifer Ryan. 2006. Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of experimental psychology: Learning, Memory, and Cognition* 32: 1341. [CrossRef]

Brito, Natalie H., and Rachel Barr. 2012. Influence of bilingualism on memory generalization during infancy. *Developmental Science* 15: 812–16. [CrossRef] [PubMed]

Brito, Natalie H., and Rachel Barr. 2014. Flexible memory retrieval in bilingual 6-month-old infants. *Developmental Psychobiology* 56: 1156–63. [CrossRef] [PubMed]

Brito, Natalie H., Amanda Grenell, and Rachel Barr. 2014. Specificity of the bilingual advantage for memory: Examining cued recall, generalization, and working memory in monolingual, bilingual, and trilingual toddlers. *Frontiers in Psychology* 5: 1369. [CrossRef] [PubMed]

Brito, Natalie H., Nuria Sebastian-Galles, and Rachel Barr. 2015. Differences in language exposure and its effects on memory generalization in monolingual, bilingual, and trilingual infants. *Bilingualism Language and Cognition* 18: 670–82. [CrossRef]

Brito, Natalie H., William P. Fifer, Dima Amso, Rachel Barr, Martha Ann Bell, Susan Calkins, Albert Flynn, Hawley E. Montgomery-Downs, Lisa M. Oakes, John E. Richards, and et al. 2019. Beyond the Bayley: Neurocognitive assessments of development during infancy and toddlerhood. *Developmental Neuropsychology* 44: 220–47. [CrossRef] [PubMed]

Brito, Natalie H., Ashley Greaves, Ana Leon-Santos, William P. Fifer, and Kimberly G. Noble. 2021. Associations between bilingualism and memory generalization during infancy: Does socioeconomic status matter? *Bilingualism: Language and Cognition* 24: 231–40. [CrossRef] [PubMed]

Carlson, Stephanie M., and Andrew N. Meltzoff. 2008. Bilingual experience and executive functioning in young children. *Developmental Science* 11: 282–98. [CrossRef]

Costa, Albert, Mireia Hernández, and Núria Sebastián-Gallés. 2008. Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition* 106: 59–86. [CrossRef]

DeAnda, Stephanie, Laura Bosch, Diane Poulin-Dubois, Pascal Zesiger, and Margaret Friend. 2016. The Language Exposure Assessment Tool: Quantifying language exposure in infants and children. *Journal of Speech Language & Hearing Research* 59: 1346–56.

Eichenbaum, Howard. 1997. Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology* 48: 547–72. [CrossRef]

Gershon, C. Richard, Karon F. Cook, Dan Mungas, Jennifer J. Manly, Jerry Slotkin, Jennifer L. Beaumont, and Sandra Weintraub. 2014. Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society* 20: 642–51. [CrossRef] [PubMed]

Gunnerud, Hilde Lowell, Dieuwer Ten Braak, Elin Kirsti Lie Reikerås, Enrica Donolato, and Monica Melby-Lervåg. 2020. Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin* 146: 1059. [CrossRef] [PubMed]

Hayne, Harlene. 2006. Age-related changes in infant memory retrieval: Implications for knowledge acquisition. *Processes of Brain and Cognitive Development: Attention and Performance* XXI: 209–31.

Karmiloff-Smith, Annette. 1994. Précis of beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences* 17: 693–745. [CrossRef]

Kids Count Data Center. 2019. Children Who Speak a Language Other than English at Home in the United States. Available online: https://datacenter.kidscount.org/data/tables/81-children-who-speak-a-language-other-than-english-at-home#detailed/1/any/false/2048,1729,37,871,870,573,869,36,868,867/any/396,397 (accessed on 28 September 2019).

Landis, J. Richard, and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159–74. [CrossRef]

Lowe, J. Cassandra, Isu Cho, Samantha F. Goldsmith, and Bruce J. Morton. 2021. The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological science* 32: 1115–46. [CrossRef]

Mahy, Caitlin E. V., and Yuko Munakata. 2015. Transitions in executive function: Insights from developmental parallels between prospective memory and cognitive flexibility. *Child Development Perspectives* 9: 128–32. [CrossRef]

Meltzoff, Andrew N. 1990. Towards a developmental cognitive science. *Annals of the New York Academy of Sciences* 608: 1–37. [CrossRef]

Riggins, Tracy, Carol L. Cheatham, Emily Stark, and Patricia J. Bauer. 2013. Elicited imitation performance at 20 months predicts memory abilities in school-aged children. *Journal of Cognition and Development* 14: 593–606. [CrossRef]

Rocha-Hidalgo, Joscelin, and Rachel Barr. 2022. Defining Bilingualism during Infancy and Toddlerhood: A Scoping Review. *International Journal of Bilingualism*. [CrossRef]

Rocha-Hidalgo, Joscelin, Mary Feller, Olivia A. Blanchfield, Sarah C. Kucker, and Rachel F. Barr. 2021. Patterns of mutual exclusivity and retention: A study of monolingual and bilingual 2-year-olds. *Infancy* 26: 1011–36. [CrossRef] [PubMed]

Rocha-Hidalgo, Joscelin, Meghana Malyala, Brianna Miller, Kate Freda, Olivia A. Blanchfield, Natalie H. Brito, and Rachel Barr. forthcoming. How language environments and vocabulary change longitudinally: An Examination of the language assessment tool at 18, 24, and 36 Months. *in prep.*.

Rose, Susan A., Judith F. Feldman, and Jeffery J. Jankowski. 2005. Recall memory in the first three years of life: A longitudinal study of preterm and term children. *Developmental Medicine & Child Neurology* 47: 653–59.

Rusnak, Sylvia N., Joscelin Rocha-Hidalgo, Olivia A. Blanchfield, Mackenzie Odier, Tuleen Sawaf, and Rachel Barr. 2022. The development of the object sequencing imitation task to measure working memory in preschoolers. *Journal of Experimental Child Psychology* 218: 105372. [CrossRef] [PubMed]

Strid, Karin, Tomas Tjus, Lars Smith, Andrew N. Meltzoff, and Mikael Heimann. 2006. Infant recall memory and communication predicts later cognitive development. *Infant Behavior and Development* 29: 545–53. [CrossRef]

Williamson, Rebecca A., Vikram K. Jaswal, and Andrew N. Meltzoff. 2010. Learning the rules: Observation and imitation of a sorting strategy by 36-month-old children. *Developmental Psychology* 46: 57–65. [CrossRef]