**Supplementary Material**

**Initial assessments**

***Method***

At the start of the study all participants were assessed with a sorting task, an attention task and two cognitive tasks to ensure that the two participant groups (shape training group and control group) did not differ in their ability to sort objects by shape and attend to stimuli, to rule out any cognitive delays, and to make sure that both groups did not differ in their visual-spatial and working memory abilities.

**Sorting Task.** This task was used to assess if participants showed a preference to organise objects by shape before the start of the study and if there were any differences between groups in their sorting preferences. Each child was presented with 16 novel objects. All objects were either red or green, were made of either foam or corrugated paper, and were of one of two possible shapes (see Figure S1). How to sort by shape was shown to them with examples. Objects were mixed and placed on a table in front of the child. Two containers were also placed on the table, one to the left and the other on the right side. The researcher started the task by saying “*We are going to play a game. We are going to put together objects that are the same”.* The researcher took one of the objects (e.g., a red round object made of corrugated paper), placed it in a container and said: *“This one goes here”*. Then, the researcher took another object with a different shape, but with the same colour and texture as the first one (e.g. a red spikey object made of corrugated paper), placed the object in a second container and said: “*And this one goes here”*. The researcher then demonstrated what the participant had to do with two more objects that also differed in shape and had the same colour and texture (e.g. a green round object made of foam and a green spiky object made of foam). Afterwards, the researcher said: *“Now it is your turn, can you help me putting the objects that are like these* (pointing to the container on the left) *here, and the objects that are like these* (pointing to the container on the right) *here?”.* Participants then sorted 12 remaining objects. The sorting task lasted approximately 5 minutes.

**Figure S1**

*Stimuli Used in the Sorting Task*

**

*Note.* All objects were either green or red, and either made of foam or corrugated paper. They were also of one of two possible shapes.

 **Attention Task.** All of the participants took part in an attention task to assess any differences in attention ability between groups. The child participant sat on their parent’s lap in front of a computer screen located at a distance of approximately 60 cm from the participant. The computer screen was connected to an eye-tracking camera (EyeLink 1000) which was located underneath the screen and which recorded the participant’s looking position and looking times. Before the start of the task, a 9 point calibration sequence was presented. During this calibration, a small happy face was shown accompanied by a *beep* on nine different sections of the monitor in a random order (left-top, centre-top, right-top, left-centre, centre, right-centre, left-bottom, centre-bottom and right-bottom). If the calibration was successful, the experiment moved on to the attention task.

For the attention task, participants watched five short video clips (35 seconds each) that consisted of a person with two objects placed on the right and left side of the video (and screen). In all videos, no object was moved for the first 10 seconds. For the following 25 seconds, the person in the video started moving one of the novel objects. The object (distracting object) on the other side of the screen was not moved (see Figure S2 for an example). All five videos followed the same procedure, but different objects were used for each video. The object that was moved and the side on which the object was being moved was counterbalanced and randomised between participants. The order of presentation of each clip was randomised between participants. A 3-second attention grabber was introduced after each video.

**Figure S2**

*Still Frames of an Example Video Used in the Attention Task*

[BLINDED FOR PEER REVIEW]

*Note.* Each video lasted 30 seconds. During the first 10 seconds no object was moved. During the following 25 seconds, the target object was moved.

**Cognitive Assessments.** During the second visit (week 2 of the study), two cognitive assessments were presented to all participants to examine if any cognitive delays were present and to make sure that both groups did not differ in their visual-spatial and working memory abilities. These two assessments were the Zoo Locations (working memory) and the Block Design subtests (i.e. visual-spatial abilities) from the Wechsler Preschool & Primary Scale of Intelligence - Fourth UK Edition (WPPSI-IV UK). The administration of the two subtests lasted around 20 minutes.

***Results***

#### **Sorting Task.** The participants in the shape training group sorted 42.85% (*SD* = 10.12) of the objects correctly by shape and participants in the control group sorted 41.66% (*SD* = 19.83) of the objects correctly by shape. In both groups, participants’ choices did not differ significantly from chance (chance being 50%) (shape training group *t*(6) = -1.86, *p* = .111; control group *t*(6) = -1.11, *p* = .308 (see Figure S3). Sorting choices did not differ significantly between groups, *t*(12) = 0.14, *p* = .890. Therefore, we did not find evidence that participants were able to sort objects by shape. However, there was also no difference between groups.

**Figure S3**

*Percentage of Correct Responses for the Two Participant Groups in the Sorting Task*

**

*Note.* The dotted line represents chance level (50%). Error bars represent 95% CIs of the means.

 **Attention Task.** Two participants from the shape training group were removed from the analyses of the attention task because they did not look at any of the videos. For the rest of the participants, the total looking times (in seconds) to all five videos of the attention task were determined (see Figure S4). To assess if both participant groups attended equally to the videos and to assess if their attention reduced equally across the trials in both groups, we analysed their total looking times (in seconds) to the first and last video. A 2 (Group: shape training group vs. control group) x 2 (Video: first video vs. fifth video) ANOVA was conducted. Participants in the shape training group looked to the first video for on average 14.45 seconds (*SD* = 12.46) and to the fifth video for 4.38 seconds (*SD* = 9.79). Participants in the control group looked for on average 21.57 seconds (*SD* =10.71) to the first video and 4.16 seconds (*SD* = 10.20) to the last video. The results of the 2 x 2 ANOVA showed a significant main effect of Video, *F*(1, 10) = 16.27, *p* = .002, ηp2 = .61, but no significant main effect of Group, *F*(1, 10) = 0.42, *p* = .530, ηp2 = .04. Results also showed no significant interaction between Group and Video, *F*(1, 10) = 1.16, *p* = .307, ηp2 = .10. Thus, the participants in both groups looked for a significantly shorter duration to the last video than to the first video, but the looking times did not differ across the two groups.

**Figure S4**

*Looking Times Across the Two Participant Groups for all Videos in the Attention Task*

**

*Note.* Error bars represent 95% CIs*.*

**Cognitive Assessments.** To rule out any significant developmental delay and to assess if both participant groups were comparable in terms of cognitive development, results from the two cognitive assessments were analysed.

***Block Design Subtest.***To assess participants’ visual-spatial abilities, scores from the block design subtest were analysed. One participant of the word training group did not want to participate in the block design task, and was therefore not included in this analysis.

Participants in the shape training group had a mean chronological age of 33.42 months and obtained scores equivalent to 33 months (*SD* = 2.44) in this task. There was a difference of -0.42 months between their chronological age and their equivalent age, however this difference was not statistically significant *t*(6) = 0.24, *p* = .811. Participants of the control group had a mean chronological age of 31.83 months and obtained scores equivalent to 33 months (*SD* = 2.00). There was a difference of +1.17 between their chronological age and their equivalent age, however this difference was not statistically significant *t*(5) = -0.56, *p* = .595. An independent samples t-test revealed that the difference between chronological age and the equivalent age was not statistically different between groups *t*(11) = 0.60, *p* = .560. Therefore, the participants in both groups showed similar visual-spatial abilities, and these abilities did not differ from what was expected given their chronological age.

***Zoo Locations Subtest.*** *T*o assess working memory, scores from the zoo locations subtest were analysed. The participants in the shape training group had a mean chronological age of 33.42 months and had scores in this subtest equivalent to 32.28 months (*SD* = 3.25). While there was an average difference of -1.14 months between their chronological age and their equivalent age, this was not statistically significant *t*(6) = .59, *p* = .571.

Participants in the control group had a mean chronological age of 31.57 months and obtained scores in this subtest equivalent to 32.14 months (*SD*= 3.48). There was an average difference of +0.57 months. between their chronological age and their equivalent age, but this was not statistically significant, *t*(6) = -.29, *p* = .780. An independent samples t-test revealed that the difference between chronological age and equivalent age was not statistically significant between groups, *t*(12) = 0.62, *p* = .542. Thus, participants in both groups had similar scores, and these scores did not differ from what it is expected given their chronological age.

**First- and second order generalisations**

**Figure S5**

*Percentage of Shape, Colour and Texture Choices in First- and Second-order Generalisation Tasks*

**

**Expressive vocabulary growth**

A further analysis was conducted including the 28 words introduced in the control session. We found no main effect of Group (*F*(1,11) = 0.65, *p* = .435 , ηp2 = .05), Testing time (*F*(1,11) = 1.13, *p* = .310 , ηp2 = .09) or Word type (*F*(1,11) = 0.00, *p* = .989 , ηp2 < .001). Similarly, there was no significant interaction between Testing time and Group, *F*(1,11) = 0.65, *p* = .435, ηp2 = .05, Word type and Group *F*(1,11) = 0.40, *p* = .537 , ηp2 = .03, and Testing time and Word type, *F*(1,11) = 3.14 *p* = .104, ηp2 = .22. The three-way interaction between Testing time, Word type and Group was not significant either, *F*(1,11) = 0.28, *p* = .604, ηp2 = .02. However, we found a significant main effect initial expressive vocabulary size, *F*(1,10) = 94.85 *p* <.001, ηp2 = .92), a significant interaction between Testing time and initial expressive vocabulary size, *F*(1,11) = 11.00, *p* = .007, ηp2 = .50) as well as a significant interaction between Word type and initial expressive vocabulary size, *F*(1,11) = 27.99, *p* < .001, ηp2 = .71). Therefore, results were similar to the analysis reported in the main manuscript that excluded the 28 labels used in the control sessions.