Title
Two-year-olds' executive functioning: The influence of task-specific vocabulary knowledge

Authors
Schonberg, Christina C
Atagi, Natsuki
Sandhofer, Catherine M

Publication Date
2018-11-01

DOI
10.1016/j.infbeh.2018.09.004

Peer reviewed

Infant Behavior and Development 53 (2018) 33–42

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Full length article

Two-year-olds' executive functioning: The influence of task-specific vocabulary knowledge

Christina C. Schonberg[a,*], Natsuki Atagi[b], Catherine M. Sandhofer[c]

[a] UW-Madison, United States
[b] UC-Riverside, United States
[c] UCLA, United States

ARTICLE INFO

Keywords:
Vocabulary
Task language
Cognitive flexibility
Preschool

ABSTRACT

Although many executive function (EF) tasks require only nonverbal responses, the language used by experimenters to explain the task may be important for young children's EF task performance. This study investigated how the vocabulary used in explaining an EF task affects 2-year-olds' performance. Experiment 1 used the standard instructions for the Reverse Categorization Task, in which children are asked to sort different-sized blocks into different-sized buckets according to one rule and then switch to a new rule. In Experiment 2, the task remained the same, but different instructions requiring less knowledge of size words were used. Children's productive vocabulary was assessed in both experiments but was only correlated with task performance in Experiment 1. These results suggest that task-specific vocabulary knowledge may play a role in children's performance on tasks designed to measure nonverbal cognitive ability.

1. Introduction

Although executive function (EF) tasks often require only nonverbal responses from participants, experimenters typically must verbally explain the task to participants in some way. In this way, the language involved in explaining the task may play a large role in EF task performance, especially for young children. Specifically, the instructions—and the language used in the instructions—as well as children's comprehension of the instructions would presumably affect children's performance on EF tasks. In this paper, we examine the role of language in a task that measures EF—specifically, rule switching.

Many studies have examined the beginnings of EF development in preverbal infants (using, for example, the antisaccade or A-not-B tasks; e.g., Bell & Adams, 1999; Johnson, 1995; Marcovitch & Zelazo, 1999) and its improvement throughout childhood (e.g., Anderson, 2010; for review, see Zelazo, Carlson, & Kesek, 2008; Zelazo & Müller, 2002), but few have examined it specifically in two-year-olds. At this age, children show wide variability in their language development and vocabulary knowledge (Fenson et al., 1994), making it an ideal time to investigate how individual differences in language contribute to performance on EF tasks. The present study thus examined the role of language in two-year-olds' performance on an EF task.

How instructions and children's comprehension of the instructions affect EF task performance have been examined with the Dimensional Change Card Sort (DCCS) task. In this task, children are presented with cards that have pictures that can either be sorted by color or by shape (e.g., cards with blue objects, half of which are rabbits and half of which are boats; cards with red objects, half of which are rabbits and half of which are boats). Children are first instructed to sort the cards by one dimension; for example, if they

* Corresponding author at: University of Wisconsin-Madison, Department of Psychology, Brogden Hall, 1202 West Johnson Street, Madison, WI 53706-1611, United States.
E-mail address: cschonberg@wisc.edu (C.C. Schonberg).

https://doi.org/10.1016/j.infbeh.2018.09.004
Received 5 February 2018; Received in revised form 25 June 2018; Accepted 17 September 2018
Available online 27 September 2018
0163-6383/

were instructed to sort by color, they would sort the red and blue cards into different boxes, regardless of whether a rabbit or boat was depicted on the card. After sorting six cards by color, children are then instructed to switch to a new sorting rule—in this case, shape (i.e., rabbits would go in one box and boats would go in the other, regardless of the color of the rabbits and boats). Three-year-olds typically have a great deal of difficulty in switching to the new rule, even though they can correctly answer the experimenter's questions about the rules (e.g., "Can you show me where the boats go?"; e.g., Perner & Lang, 2002; Zelazo, Frye, & Rapus, 1996). Interestingly, performance on the DCCS is improved when children are instructed to state the rules of the task and label the relevant dimension themselves prior to sorting each card, in contrast to the experimenter reminding them of the rules (Kirkham, Cruess, & Diamond, 2003; Ramscar, Dye, Gustafson, & Klein, 2013; Towse, Redbond, Houston-Price, & Cook, 2000; see also Miller & Markovitch, 2011, for the benefits of labeling in a different EF task). These findings suggest that children's performance on an EF task improves when children are able to restate the instructions; this may be driven by the fact that they have the language skills to understand and produce the instructions themselves. Moreover, these findings suggest that the instructions used to explain EF tasks may be even more important for two-year-olds because their language skills are less advanced.

In preschoolers, language development has been found to longitudinally predict performance on EF tasks (e.g., Fuhs & Day, 2011; Kuhn, Willoughby, Wilbourn, Vernon-Feagans, & Blair, 2014). Within the same school year, preschoolers' language skills at the beginning of the year have been found to predict their EF skills at the end of the year (Fuhs & Day, 2011). Language development predicts EF development on a longer time scale as well: the *rate* of a child's change in language between 15 and 36 months is predictive of their *rate* of change in EF between 36 and 60 months, as well as their EF at 60 months (Kuhn, Willoughby, Vernon-Feagans, & Blair, 2016). This predictive relation could be due in part to the language demands of EF tasks: children with more advanced language development could simply be better able to understand what they are being asked to do.

Moreover, many studies have also shown that language affects performance in EF tasks even in older children and adults (Fatzer & Roebers, 2012; Karbach & Kray, 2007; Kray, Eber, & Karbach, 2008; McNamara & Scott, 2001; Miyake, Emerson, Padilla, & Ahn, 2004; Zelazo & Frye, 1998; Zelazo, Reznick, & Piñon, 1995; see Cragg & Nation, 2010, for a review). For adults, performance on selective attention (i.e., visual search) tasks improves when they are instructed to say the name of the target item out loud (Lupyan & Swingley, 2012). In contrast, both older children and adults show a decline in performance on memory-demanding tasks (e.g., remembering complex spans of stimuli) or tasks involving switching when they simultaneously engage in an articulatory suppression task (e.g., repeating the syllable 'da'; Fatzer & Roebers, 2012; Kray et al., 2008). It is thought that children typically use inner speech to scaffold their working memory and succeed in the task; the articulatory suppression task prevents the use of inner speech, and performance is thus negatively affected. Overall, these results are consistent with the idea that language helps with different aspects of EF across development.

There are different theoretical explanations as to why EF tasks are difficult for young children. Cragg and Nation (2010) highlight the importance of inner speech; Kirkham et al. (2003) propose 'attentional inertia'—that is, that poor EF task performance results from having difficulty in disengaging one's attention from a previously-relevant feature or stimulus; and Miller and Marcovitch (2011) use the framework of the Hierarchical Competing Systems Model, which emphasizes the idea that repeated actions create a habit-based response. These theoretical perspectives focus on the role of cognitive processes and attentional focus in children's EF task performance, but they presuppose that the child understands what they are being asked to do in the task. Thus, by examining the role of language in EF tasks, this study investigates the underlying assumption in these theories that children understand task instructions.

The studies described above find that language production (or the prevention of language production) influences performance on EF tasks throughout childhood. Because the EF tasks used with young children almost always involve verbal instructions to the child as to what to do, the present study asks the novel question of whether the phrasing of task instructions—as well as children's language skills—are related to EF task performance when the task itself remains the same. To do so, we used two common types of EF instructions in a rule-switching task and measured children's knowledge of task-relevant words. Although EF tasks themselves are generally nonlinguistic in nature, language skills are required in order to understand the instructions and succeed in the tasks.

In Experiment 1, we examined how children's vocabulary (specifically, knowledge of task-relevant words) relates to task performance. We collected data from two-year-olds using the Reverse Categorization Task (RCT; Carlson, Mandell, & Williams, 2004) because there is a great deal of variability both in word knowledge and RCT performance at this age. In Experiment 2, we changed the language used in the task instructions by adapting the DCCS protocol to investigate how this change affects performance. Although the DCCS is a task intended for older children, the published protocol uses fewer dimension-relevant words in the task instructions and thus has fewer vocabulary demands (Zelazo, 2006). We adapted this protocol to use size words (i.e., "big" and "little") instead of color and shape words. We also introduced a behavioral measure of receptive vocabulary. In both experiments, the child was instructed to do the same thing: to first sort small blocks into a small bucket and large blocks into a large bucket, and then switch the sorting rule (i.e., sort small blocks into a large bucket and large blocks into a small bucket). The results of these experiments highlight the importance of the language that is used in EF tasks, particularly in preschoolers who have variable and limited language skills.

2. Experiment 1

In this experiment, we asked how two-year-olds' vocabulary (as measured by the MCDI) related to their performance on the RCT, a measure of EF commonly used at this age. Specifically, we were interested in their knowledge of the size words used in the standard RCT instructions.

Fig. 1. Wooden blocks and buckets that were used in Experiment 1 and Experiment 2.

2.1. Method

2.1.1. Participants

Twenty-seven children between the ages of 24–30 months (M_{age} = 27.17 months, SD = 1.72 months) participated in this study. Participants were recruited through local preschools and through a shared participant database at the researchers' university. Protocols were approved by the university's institutional review board and parents provided informed consent prior to the child's participation. All children were learning English as their primary language; the MacArthur-Bates Communicative Development Inventory (MCDI) was used to assess English proficiency. An additional six children were tested but excluded from the primary analysis due to their failure to learn the rule during the preswitch phase, which was defined as requiring correction from the experimenter on more than one preswitch trial. Their data are analyzed separately below.

2.1.2. Reverse categorization task

2.1.2.1. Materials. We modeled this task after the RCT used in Carlson et al. (2004). Twelve wooden blocks (six 2" × 2" × 2" blocks and six 1" × 1" × 1" blocks) of identical color, texture, and material were used in this task; the only dimension in which the blocks differed was size. A small wooden bucket (5" high, 6.5" in diameter) and a larger wooden bucket (7.5" high, 11.5" in diameter) were also used in the task (see Fig. 1).

2.1.2.2. Procedure. Each child was tested individually, seated at a small table across from the experimenter. The task consisted of three phases, always in the same order: the training phase (six trials), the preswitch phase (six trials), and the postswitch phase (12 trials).

2.1.2.2.1. Training phase. In the training phase, the experimenter placed the two buckets on the table and labeled them by size (e.g., "I have a big bucket and a little bucket."), then told the child that they were going to play a game in which "the big blocks go in the big bucket and the little blocks go in the little bucket." The experimenter then demonstrated sorting big blocks (N = 3) and little blocks (N = 3), alternating between block sizes and saying the rule each time they showed the child a block and placed it in the bucket (i.e., "This is a big block. It goes in the big bucket," or "This is a little block. It goes in the little bucket.").

2.1.2.2.2. Preswitch phase. After sorting the first six blocks, the experimenter then told the child it was their turn to sort. The experimenter held a block between the two buckets for the child to see, stated the size of the block (e.g., "This is a little block.") and the corresponding sorting rule (e.g., "It goes in the little bucket."), and placed the block on the table equidistant between the two buckets. The child then placed the block in one of the buckets. If the child made an error, the experimenter corrected them (see Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011), by taking the block out and repeating the size and rule (e.g., "This is a little block. It goes in the little bucket."), until they sorted the block correctly. Children who were corrected more than once (N = 6) were excluded from data analysis. This phase of the task concluded after the child sorted all six blocks. The experimenter then took the buckets off of the table and, behind the table where the child could not see, removed the blocks from the buckets to prepare for the next phase.

2.1.2.3. Postswitch phase. In the postswitch phase of the task, the experimenter told the child that they were going to play a new game (the "silly game") and then placed the buckets back on the table. The experimenter told the child that in the "silly" game, "the big blocks go the little bucket and the little blocks go in the big bucket." Following Carlson et al. (2004), there was no demonstration of the new rule; the experimenter simply held a block equidistant between the two buckets for the child to see, labeled its size (e.g., "This is a big block."), and stated the rule (e.g., "It goes in the little bucket."). After the child had placed the block in a bucket, the experimenter made a neutral statement (e.g., "Let's do another one!") and presented the child with the next block. All 12 trials proceeded in this manner.

2.2. Results and discussion

Because children's preswitch performance was corrected and children were excluded if they were corrected more than once during the preswitch phase, all children sorted 6 out of 6 preswitch trials correctly. In our statistical analyses, we first asked whether children would be able to successfully switch to the new rule in the postswitch phase of the task. We defined postswitch performance as the proportion of blocks correctly sorted (out of 12). A one-sample *t*-test revealed that as a group, children's postswitch performance ($M = 0.308$, $SD = 0.402$) was significantly below chance (0.50), $t(26) = 2.468$, $p = .020$; this finding indicates that, overall, children perseverated during the postswitch phase by using the preswitch rule. We analyzed individual children's performance using the binomial distribution and found a similar pattern of results. Assuming a chance probability of success of 0.5, across 12 trials, the probability of two or fewer successes is $p = .0193$, and the probability of 10 or more successes is also $p = .0193$. Thus, the total probability of 0, 1, 2, 10, 11, or 12 successes is quite small, below 0.05. Comparing each child's performance to this distribution, we found that 17 children perseverated (i.e., responded correctly on two or fewer trials, $p < .05$), four children performed at chance (i.e., neither perseverated nor switched, $p > .05$), and six children successfully switched (i.e., responded correctly on 10 or more trials, $p < .05$).

We then conducted analyses to determine how vocabulary size was related to this outcome. There was a great deal of variability in both postswitch performance (range: 0.00–1.00) and vocabulary size ($M = 484.15$ words, $SD = 172.09$ words; range: 178–757 words). Overall vocabulary size was found to be significantly correlated with postswitch performance, $r(25) = .539$, $p = .004$. Furthermore, we found that knowledge of size words on the MCDI (i.e., "big," "little," "long," "tiny") was significantly correlated with postswitch performance, $r(25) = .491$, $p = .011$.

One possibility is that children with larger vocabularies knew more size words than children with smaller vocabularies, and thus knowing size words specifically was helpful in a task in which the instructions included the words "big" and "little." To examine the effects of specifically knowing the size words used in the task, we used children's MCDIs to classify them into one of three groups: those who produced both "big" and "little" ($N = 15$), those who produced only one of the task-relevant words (i.e., "big" *or* "little"; $N = 8$), or those who produced neither word ($N = 3$). Due to issues of statistical power that would arise if the smaller groups were analyzed separately, children who produced zero or one of the task-relevant words were combined into one group for analysis. An independent-samples *t*-test showed that children who produced both size words had marginally higher performance on the postswitch trials ($M = 0.400$, $SD = 0.438$) than children who only produced zero or one of the size words ($M = 0.121$, $SD = 0.231$), $t(21) = 1.914$, $p = .068$ (see Fig. 2). Knowing both "big" and "little" is important for success in this particular task because the instructions in the postswitch phase explicitly state where the block should go ("This is a little block. It goes in the big bucket."). If the child does not understand the meanings of both "little" and "big," it would be difficult for them to understand that these instructions are different from what they were told in the preswitch phase. Moreover, "big" and "little" are words that describe relative, not absolute, sizes of objects. Therefore, it may be that knowing only one of these words is not enough to understand that the word describes a relational characteristic. Rather, it may be necessary to know *both* of these words in order to understand how they map onto the objects they describe.

An alternative possible explanation for the relation between vocabulary and postswitch performance is that as children's age increases, both their vocabulary and cognitive skills increase. Therefore, vocabulary could be serving as a proxy for age in this instance. However, there was no relation between age and postswitch performance, $r(25) = 0.138$, $p = .493$. Our results suggest that performance in an EF task such as the RCT may be influenced in part by the participant's understanding of the words used in the task (see Jacques & Zelazo, 2005, for review).

2.2.1. Excluded children

We separately analyzed data from children who were excluded from the main analysis due to the fact that they did not pass the preswitch phase (i.e., had to be corrected more than once; $N = 6$). Because of the small sample size, no inferential statistics investigating group performance were conducted. As described above, we used the binomial distribution to examine individuals' performance during both preswitch and postswitch trials. In each of the six preswitch trials, the chance probability of success is 0.5;

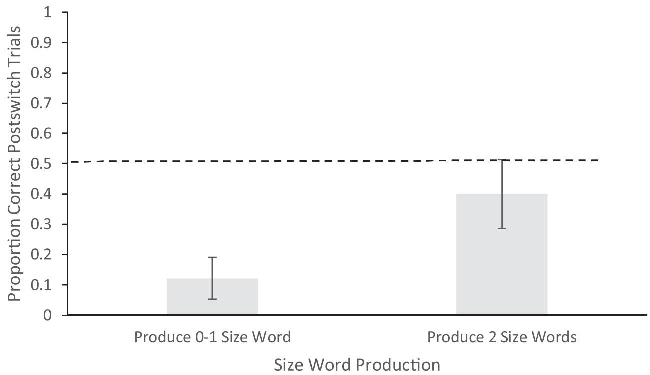

Fig. 2. Children's postswitch performance in Experiment 1. Error bars represent standard error; the dashed line indicates chance performance.

the probability of one success is $p = .016$, and the probability of six successes is also $p = .016$. Here, we found that each child performed no differently from chance during this phase (i.e., each child had between 2–5 successes, $p > .05$). In the postswitch phase, which had 12 trials, the same criteria as described above were used. Two children correctly sorted two or fewer blocks, $p < .05$, and four children performed at chance, $p > .05$. Task-relevant word knowledge, as assessed by the MCDI, was similar to that of the children in the main analysis; here, two children produced one of the task-relevant words, and four children produced both words ("big" and "little"). These results suggest that other factors aside from vocabulary knowledge (e.g., inattention) may have contributed to these children's failure to pass the preswitch phase. In addition, the MCDI does not necessarily capture children's understanding of the relational nature of the words "big" and "little"; it is possible that a child could say the word "big" (e.g., "I'm a big girl!") without understanding its meaning relative to the word "little". Therefore, in Experiment 2, we introduce a task that addresses children's understanding of these relational terms.

3. Experiment 2

In Experiment 2, we modified the task instructions from Experiment 1, following the instruction format of the DCCS (Zelazo, 2006). The DCCS and our corresponding modification use fewer dimension-specific labels during the sorting task (e.g., the size of the block is labeled in each trial, but the size of the bucket is not). Because we found in Experiment 1 that knowledge of specific words used in the RCT was strongly related to performance on the task, we hypothesized in this experiment that if children were given instructions that contained fewer dimension-specific words, more children may successfully switch rules than in Experiment 1. However, if this particular block-sorting task consistently measures the same construct, independent of the instructions that are used, we would expect children's pattern of performance in this experiment to not differ from Experiment 1.

The DCCS typically uses two relevant dimensions (shape and color) for a sorting task; in our modification, we only used one dimension (size) in order to make it developmentally appropriate for 24- to 30-month-olds and changed the instruction script (Zelazo, 2006) accordingly. In addition, we added a behavioral measure of the child's understanding of the words "big" and "little." This measure allowed us to assess the child's comprehension knowledge of these words in a more in-depth manner than the MCDI, a parent-report measure of production, could provide.

3.1. Method

3.1.1. Participants

Twenty-three children between the ages of 24 and 30 months ($M_{age} = 26.89$ months, $SD = 1.74$ months) participated in this study. As in Experiment 1, parents provided informed consent prior to their child's participation. Participants were recruited and assessed for English proficiency using the same methods described in Experiment 1 (i.e., the MCDI); no children were excluded.

3.1.2. Tasks

3.1.2.1. One-dimensional sorting task. In Experiment 2, the same wooden blocks and buckets as Experiment 1 were used. As in Experiment 1, each child was tested sitting across from the experimenter at a small table. This task consisted of three phases, always in the same order: the training phase (two trials), the preswitch phase (10 trials), and the postswitch phase (12 trials).

3.1.2.1.1. Training phase. In the training phase, the experimenter introduced the game and labeled the blocks and buckets by size, saying, "Here's a big bucket and here's a little bucket. Now, we're going to play a game. This is the 'same' game. In the 'same' game, all the little blocks go here [pointing to the little bucket], and all the big ones go there [pointing to the big bucket]." The experimenter then held up one block, equidistant between the two buckets. The experimenter labeled and sorted the block (e.g., "See, here's a big one. So it goes here."), repeating the preswitch rules afterward ("If it's big it goes here, but if it's little it goes there."). Next, the experimenter presented the child with a block of the other size, holding the block equidistant between the two buckets, and asked the child to sort the block ("Now here's a little one. Where does this one go?"). If the child made a mistake, the experimenter took the block out of the bucket and corrected the child (e.g., "No, this one's little, so it has to go over here in the 'same' game. Can you help me put this little one down?"), ensuring that the child sorted the block correctly after the correction.

3.1.2.1.2. Preswitch phase. Immediately after the two training trials, the experimenter moved into the preswitch phase of the task, saying to the child, "Now it's your turn. So remember, if it's big it goes here, but if it's little it goes there." The experimenter then presented a block to the child, labeled it by size (e.g., "Here's a little one."), placed it on the table equidistant between the two buckets, and asked the child to sort it ("Where does it go?"). Regardless of whether the child sorted the block correctly, the experimenter made a neutral statement (e.g., "Let's do it again!") and presented the child with the next block. In total, the child sorted 10 blocks in the preswitch phase.

3.1.2.1.3. Postswitch phase. After the child had sorted all the blocks in the preswitch phase, the post-switch phase began. The experimenter removed the buckets from the table, emptied the blocks from the buckets (out of the child's view) and said to the child, "Now we're going to play a new game. We're not going to play the 'same' game anymore. We're going to play the 'different' game." After placing the empty buckets back on the table, the experimenter said, "In the 'different' game, all the big blocks go here [pointing to the little bucket], and all the little blocks go there [pointing to the big bucket]. Remember, if it's a big block, put it here, but if it's a little block put it there. Okay?"

As in the preswitch phase, the experimenter presented a block to the child, labeled it by size (e.g., "This is a big one."), and asked the child to sort it ("Where does it go?"). The experimenter did not give the child any feedback, making only neutral statements such as "Let's do another one!", and the task concluded when the child had sorted 12 blocks.

3.1.2.2. Size understanding task. This task was designed to measure children's understanding of the words "big" and "little." Sixteen sheets of paper, each with a pair of pictures, were used in this task. The pictures were black-and-white line drawings of concrete nouns (e.g., strawberry, ball). Each pair was printed on one sheet of paper such that there was either a large size difference (400%; eight pairs) or a small size difference (200%; eight pairs) between the pictures, and the pictures either showed a typical size difference (e.g., small mouse and large elephant) or an atypical size difference (e.g., small airplane and large car). The order of the picture pairs was randomized across children.

So as not to prime the children to think about size on the One-Dimensional Sorting Task, the Size Understanding Task was always administered after the One-Dimensional Sorting Task. The experimenter introduced the task by saying they were going to play a game in which the child would put a sticker either on a big picture or on a little picture. First, there were two training trials, in which the experimenter helped the child put the sticker on the requested picture. In one training trial, the experimenter requested the big picture, and in the other training trial, the experimenter requested the little picture; the request order was randomized across children. After the two training trials, the experimenter said "Now it's your turn! I'm going to give you a sticker and tell you which picture to put it on." The experimenter placed one of the picture pair sheets on the table in front of the child, said, "Can you please put the sticker on the [big/little] picture?" and gave the child a sticker. After the child placed the sticker on the picture, the experimenter made a neutral statement (e.g., "Let's do some more!") and moved on to the next trial until the child had placed a sticker on all 14 remaining picture pair sheets.

3.2. Results and discussion

In this experiment, we investigated whether there was a relation between children's performance on a one-dimensional sorting task and their knowledge of the specific dimensional words used in the task (i.e., "big" and "little"). We found that in the preswitch phase, although we did not correct sorting errors, group performance was above chance (chance = 0.50; $t(22) = 4.655$, $p < .001$; $M_{sorting\ accuracy} = 0.757$, $SD = 0.264$). Because we did not exclude children based on preswitch performance, we separately analyzed children who passed preswitch (i.e., sorted 10 out of 10 blocks correctly; N = 12) and children who did not (i.e., sorted fewer than 10 out of 10 blocks correctly; N = 11). The mean preswitch performance for children who did not sort 10 out of 10 blocks correctly was no different from chance ($M = 0.491$, $SD = 0.07$), $t(10) = 0.340$, $p = .676$.

We then used the binomial distribution to examine *individual* performance in the preswitch trials. Assuming a chance probability of success of 0.5, across 10 trials, the probability one or fewer successes is $p = .010$, and the probability of nine or more successes is also $p = .010$. Thus, the total probability of 0, 1, 9, or 10 successes is quite small, below 0.05. Comparing each child's performance to this distribution, we found that 11 children performed no differently from chance, $p > .05$, and 12 children learned the rule (i.e., responded correctly on nine or more trials, $p < .05$); this is consistent with the group-based finding reported in the previous paragraph.

As in Experiment 1, we defined postswitch performance as the proportion of blocks correctly sorted (out of 12). A one-sample *t*-test revealed that as a group, children's postswitch performance ($M = 0.587$, $SD = 0.322$) was not significantly different from chance (0.50), $t(22) = 1.295$, $p = .209$. We also separately analyzed the performance of children who passed preswitch and those who did not. For the children who passed preswitch, their postswitch performance was not significantly different from chance ($M = 0.667$, $SD = 0.431$), $t(11) = 1.341$, $p = .207$. Similarly, performance of children who did not pass preswitch was also not significantly different from chance ($M = 0.500$, $SD = 0.091$), $t(10) = 0.002$, $p = .998$. Because the variance in postswitch performance between these two groups was heterogeneous, we compared their postswitch performance using nonparametric statistics and found that the children who passed preswitch performed marginally significantly better than children who did not pass preswitch, Mann-Whitney $U = 38.500$, $p = .082$ (2-tailed). The median postswitch score for children who passed preswitch (N = 12) was 0.917, and the median postswitch score for children who did not pass preswitch (N = 11) was 0.500. Although not statistically significant at $p < .05$, children who passed preswitch had higher mean and median postswitch performance than children who did not pass preswitch. This finding shows that the overall group mean postswitch performance was not raised to chance levels by including children who did not pass the preswitch phase.

Using the binomial distribution as described for the postswitch phase of Experiment 1 (12 test trials with a 0.5 chance of success on each trial), we analyzed individual children's performance and found that three children perseverated (i.e., responded correctly on two or fewer trials, $p < .05$), 12 children performed no differently from chance ($p > .05$), and seven children successfully switched (i.e., responded correctly on ten or more trials, $p < .05$). Importantly, all children who perseverated or successfully switched in the postswitch phase had had nine or more successes (i.e., learned the original rule) during the preswitch phase.

We then conducted analyses to determine how vocabulary size, specific word knowledge, and performance on the Size Understanding Task were related to this outcome. Here, vocabulary size was not found to be significantly correlated with postswitch performance, $r(21) = 0.353$, $p = .098$.

To examine the effects of specifically knowing the size words used in the task, we analyzed data from two different sources: the MCDIs and the Size Understanding Task. As in Experiment 1, we used children's MCDIs to classify them as either producing both "big" and "little" ($N = 10$) or producing one or neither of these words ($N = 13$) Postswitch performance was analyzed using an independent-samples *t*-test, which showed that performance of children who produced both words ($M = 0.742$, $SD = 0.341$) was higher than that of children who produced one or neither of the words ($M = 0.468$, $SD = 0.260$), $t(21) = 2.187$, $p = .040$ (see Fig. 3). As in Experiment 1, understanding the relative meanings of "big" and "little" still matters for performance in both the preswitch and postswitch phases of this task. Of the 12 children who passed preswitch, eight children knew both "big" and "little"; of the 11 children who did not pass preswitch, only two children knew both of these size words. In the postswitch phase, as detailed above, a high

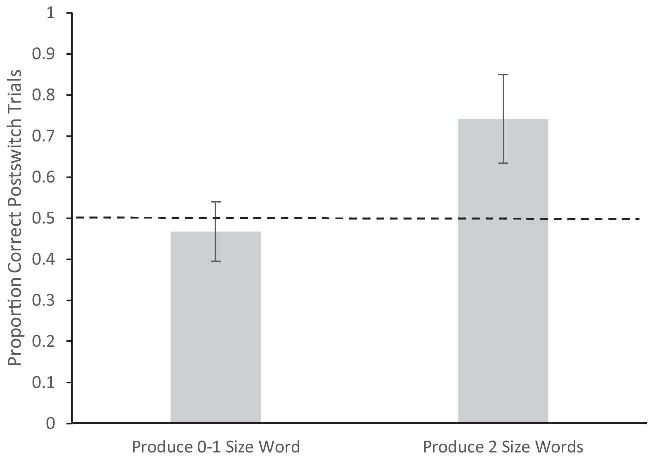

Fig. 3. Children's postswitch performance in Experiment 2. Error bars represent standard error; the dashed line indicates chance performance.

proportion of children who learned the preswitch rule successfully switched to the new rule. These findings are consistent with our prediction that the use of fewer size words in the task instructions would facilitate switching performance in this task.

Initial *t*-tests examining performance on the Size Understanding Task showed that there were no differences in performance based on magnitude of size difference (200% vs. 400%) or typicality of size difference, *p*s > .05. Overall, children performed significantly above chance (i.e., 7 out of 14 trials correct) on the task (M = 9.78, SD = 2.89), $t(22) = 4.616$, $p < .001$. Furthermore, children who produced both task-relevant words (i.e., "big" and "little"), as indicated by the MCDI, performed better on the Size Understanding Task (M = 11.2, SD = 2.86; N = 10) than those who produced one or neither of the words (M = 8.69, SD = 2.50; N = 13), $t(21) = 2.243$, $p = .036$. In the context of this experiment, we were most interested in how children's overall performance on the Size Understanding Task related to their postswitch performance on the One-Dimensional Sorting Task. We found a significant correlation between performance on these two tasks, $r(21) = 0.473$, $p = .023$. These results suggest that knowledge of *both* size words (i.e., "big" and "little") is necessary for children to successfully complete tasks in which this relative size difference is a key part of the task.

3.3. Comparing Experiments 1 and 2

Finally, we compared postswitch performance between Experiment 1 and Experiment 2. There were no significant differences in age, $t(49) = 0.507$, $p = .615$, or vocabulary size, $t(48) = 0.870$, $p = .389$, between the groups of children in each experiment. Overall, the children who participated in Experiment 2 performed better in the postswitch phase compared to the children in Experiment 1, $t(49) = 2.649$, $p = .011$ (see Fig. 4). These results corroborate our correlational findings in that parent-reported vocabulary knowledge was less related to task performance in this experiment than in Experiment 1. Furthermore, these results suggest that the use of fewer size words in the task instructions may have indeed decreased vocabulary demands on the participants in this experiment.

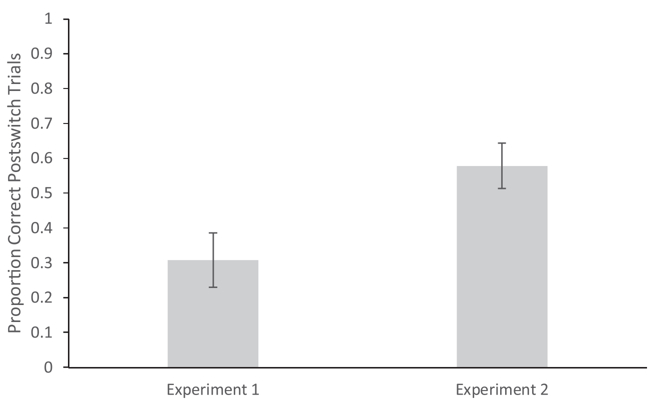

Fig. 4. Children's postswitch performance in Experiment 1 and Experiment 2. Error bars represent standard error; the dashed line indicates chance performance.

Table 1
Procedural differences between Experiment 1 and Experiment 2.

	Experiment 1	Experiment 2
Training phase		
Size of bucket stated?	Y	Y
Number of demo trials	6	2
Sorting rule stated?	Y	Y
Size of block stated?	Y	Y
Preswitch phase		
Number of child trials	6	10
Mistakes corrected?	Y	N
Sorting rule stated?	Y	N
Size of block stated?	Y	Y
Postswitch phase		
Number of demo trials	0	0
Number of child trials	12	12
Sorting rule stated?	Y	N
Size of block stated?	Y	Y

4. General discussion

In this paper, we investigated the effects of task instructions on switching performance in preschoolers. Here, we used the same task in two different experiments (i.e., sorting blocks into buckets based on a size rule that changes), modifying the task instructions in Experiment 2 to mirror a task that has been commonly used to assess cognitive flexibility. The results of Experiment 1, which used instructions that included many size words, showed that children who knew the size words used in the instructions had marginally higher postswitch performance than children who did not know both size words used in the instructions. However, the majority of children perseverated in the postswitch phase and did not switch to the new rule. In Experiment 2, which used instructions that had fewer size words, children who knew both size words had higher performance than those who did not know both words. However, of the children who performed above chance during the preswitch phase, a larger proportion of individual children successfully switched to the new rule in the postswitch phase (i.e., 22% of children who learned the preswitch rule successfully switched in Experiment 1; 58% of children who learned the preswitch rule successfully switched in Experiment 2). Taken in conjunction with our finding that there was a significant positive correlation between postswitch performance and performance on the Size Understanding Task, these results suggest that performance is related to understanding the relative nature of the task-relevant words (which is not necessarily measured by the MCDI).

Because we used two preexisting sets of instructions (i.e., the RCT in Experiment 1 and a modified DCCS in Experiment 2; Carlson et al., 2004; Zelazo, 2006), there were a few key differences between the tasks (see Table 1). These differences occurred because we used existing protocols for two commonly-used tasks—the RCT and the DCCS—that assess children's cognitive flexibility. For example, in Experiment 1, the experimenter demonstrated six trials in the training phase, whereas in Experiment 2, the experimenter demonstrated only two training phase trials. Additionally, children's mistakes were corrected during the preswitch phase of Experiment 1 (based on the RCT protocol), whereas mistakes were not corrected during the preswitch phase of Experiment 2 (based on the DCCS protocol). However, based on these differences, Experiment 2 should have been harder overall for the children: there were fewer trials demonstrated by the experimenter, they had more opportunities to build up a prepotent response in the preswitch phase (but see Marcovitch & Zelazo, 2006), and the sorting rule was only stated before the start of each phase (as opposed to before each trial, as in Experiment 1).

Nevertheless, mean postswitch performance was higher and a greater number of individual children successfully switched rules in Experiment 2 than in Experiment 1. We posit that, in addition to the differences described above, differences in the task instructions also contributed to children's differential performance in these experiments. First, in Experiment 1, the experimenter stated the sorting rule for each of the preswitch and postswitch trials; in contrast, in Experiment 2, the experimenter only stated the sorting rule at the beginning of each phase (i.e., the preswitch phase and the postswitch phase). Second, in Experiment 1, the experimenter labeled each bucket by its size whenever it was referenced (e.g., "This is a little block. It goes in the little bucket."). In contrast, in Experiment 2, the experimenter labeled each bucket by size only once, at the beginning of the task (i.e., "Here's a big bucket and here's a little bucket."). When the sorting rules were stated, the experimenter did not state the size of the bucket; they only stated the size of the block (e.g., "In the 'same' game, all the little blocks go here."; "If it's big it goes here, but if it's little it goes there."). Because the experimenter in Experiment 2 always used size words to describe the blocks, but rarely used size words to describe the buckets, fewer size words were used overall in Experiment 2.

In our analyses of Experiment 2, we found a significant positive correlation between performance on the postswitch phase of the sorting task and performance on the Size Understanding Task. This finding shows that receptive knowledge of task-relevant vocabulary was related to task performance. The use of fewer size words in Experiment 2 could have allowed children to better focus their attention on the blocks to be sorted, as these were the only objects that were consistently labeled. This is consistent with previous findings that preschoolers' EF task performance improves when the experimenter uses only task-relevant labels (Doebel & Zelazo, 2013). Relative to Experiment 2, Experiment 1 may have placed more demands on children's attention and working memory because both the blocks and buckets were consistently labeled in this experiment; high attentional demands often underlie 3-year-olds' poor performance on EF tasks, and decreasing these demands improves their performance (e.g., Perner & Lang, 2002; Rennie, Bull, & Diamond, 2004).

The question of the *extent* to which task language influenced task performance, relative to other procedural characteristics, remains open for discussion. Future investigations of the importance of task language should systematically manipulate these procedural characteristics that varied between our experiments because we used preexisting protocols. For example, it could be that an experimenter's modeling of a sorting rule impedes children's ability to switch to a new rule; in Experiment 1, the experimenter modeled six preswitch trials, whereas in Experiment 2, the experimenter modeled only two preswitch trials.

4.1. Conclusions

Through the experiments described in this report, we have shown that task language is an important characteristic of tasks that require a nonverbal response, and it may play a key role in assessments of both switching ability and broader EF constructs. Our results are consistent with work showing that performance on EF tasks designed to measure the same construct—even tasks that ask children to do the exact same thing—yield different results depending on how the task is presented (e.g., Kirkham et al., 2003; Munakata & Yerys, 2001) or which tasks are used (e.g., Deák & Wiseheart, 2015). We extend the findings of previous research by showing that even when the exact same switching task is used, the language the experimenter uses to instruct the child can impact task performance.

Our results suggest that even nonverbal tasks can have verbal demands, which raises the question of whether a truly nonverbal measure of EF can exist. In the same way that researchers have developed variations of the same EF task to account for developmental changes in cognitive capacity, attention, and memory (e.g., different versions of the DCCS and Stroop tasks; Archibald & Kerns, 1999; Berlin & Bohlin, 2002; Carlson, 2005; Gerstadt, Hong, & Diamond, 1994; Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Kochanska, Murray, & Harlan, 2000), future designers of tasks to measure nonverbal cognitive abilities should be aware of the language demands involved in the tasks and carefully consider the language they use in explaining the tasks to participants.

Funding

This work was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-0707424 to Natsuki Atagi.

Declarations of interest

None.

Acknowledgments

We thank Caitlin Choy and other members of the UCLA Language and Cognitive Development Lab for their assistance with this project. We are especially thankful to our infant participants and their families for their contributions to this research.

References

Anderson, P. J. (2010). Towards a developmental model of executive function. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.). *Executive functions and the frontal lobes* (pp. 37–56). New York: Psychology Press.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5*(2), 115–129. https://doi.org/10.1076/chin.5.2.115.3167.

Bell, M. A., & Adams, S. E. (1999). Comparable performance on looking and reaching versions of the A-not-B task at 8 months of age. *Infant Behavior and Development, 22*(2), 221–235. https://doi.org/10.1016/S0163-6383(99)00010-7.

Berlin, L., & Bohlin, G. (2002). Response inhibition, hyperactivity, and conduct problems among preschool children. *Journal of Clinical Child & Adolescent Psychology, 31*(2), 242–251. https://doi.org/10.1207/S15374424JCCP3102_09.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40*(6), 1105–1122. https://doi.org/10.1037/0012-1649.40.6.1105.

Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science, 2*(4), 631–642. https://doi.org/10.1111/j.1756-8765.2009.01080.x.

Deák, G. O., & Wiseheart, M. (2015). Cognitive flexibility in young children: General or task-specific capacity? *Journal of Experimental Child Psychology, 138*, 31–53. https://doi.org/10.1016/j.jecp.2015.04.003.

Doebel, S., & Zelazo, P. D. (2013). Bottom-up and top-down dynamics in young children's executive function: Labels aid 3-year-olds' performance on the Dimensional Change Card Sort. *Cognitive Development, 28*(3), 222–232. https://doi.org/10.1016/j.cogdev.2012.12.001.

Fatzer, S. T., & Roebers, C. M. (2012). Language and executive functions: The effect of articulatory suppression on executive functioning in children. *Journal of Cognition and Development, 13*(4), 454–472. https://doi.org/10.1080/15248372.2011.608322.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), https://doi.org/10.2307/1166093.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology, 47*(2), 404–416. https://doi.org/10.1037/a0021065.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 312–7 years old on a stroop- like day-night

test. *Cognition, 53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-X.
Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28*(2), 617–644. https://doi.org/10.1207/s15326942dn2802_4.
Jacques, S., & Zelazo, P. D. (2005). On the possible roots of cognitive flexibility. In B. D. Horner, & C. S. Tamis-LeMonda (Eds.). *The development of social cognition and communication* (pp. 53–81). New York: Psychology Press.
Johnson, M. H. (1995). The inhibition of automatic saccades in early infancy. *Developmental Psychobiology, 28*(5), 281–291. https://doi.org/10.1002/dev.420280504.
Karbach, J., & Kray, J. (2007). Developmental changes in switching between mental task sets: The influence of verbal labeling in childhood. *Journal of Cognition and Development, 8*(2), 205–236. https://doi.org/10.1080/15248370701202430.
Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449–467. https://doi.org/10.1111/1467-7687.00300.
Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*(2), 220–232. https://doi.org/10.1037/0012-1649.36.2.220.
Kray, J., Eber, J., & Karbach, J. (2008). Verbal self-instructions in task switching: a compensatory tool for action-control deficits in childhood and old age? *Developmental Science, 11*(2), 223–236. https://doi.org/10.1111/j.1467-7687.2008.00673.x.
Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., & Blair, C. B. (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *Journal of Experimental Child Psychology, 148*, 20–34. https://doi.org/10.1016/j.jecp.2016.03.008.
Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., & Blair, C. B. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development, 85*(5), 1898–1914. https://doi.org/10.1111/cdev.12249.
Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology, 65*(6), 1068–1085. https://doi.org/10.1080/17470218.2011.647039.
Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70*(6), 1297–1313. https://doi.org/10.1111/1467-8624.00095.
Marcovitch, S., & Zelazo, P. D. (2006). The influence of number of A trials on 2-year-olds' behavior in two A-not-B-type search tasks: A test of the hierarchical competing systems model. *Journal of Cognition and Development, 7*(4), 477–501. https://doi.org/10.1207/s15327647jcd0704_3.
McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10–17. https://doi.org/10.3758/BF03195736.
Miller, S. E., & Marcovitch, S. (2011). Toddlers benefit from labeling on an executive function search task. *Journal of Experimental Child Psychology, 108*(3), 580–592. https://doi.org/10.1016/j.jecp.2010.10.008.
Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*(2–3), 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004.
Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*(4), 335–337. https://doi.org/10.1111/1467-9280.00361.
Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant and Child Development, 11*(2), 93–105. https://doi.org/10.1002/icd.299.
Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology, 108*(3), 567–579. https://doi.org/10.1016/j.jecp.2010.10.009.
Ramscar, M., Dye, M., Gustafson, J. W., & Klein, J. (2013). Dual routes to cognitive flexibility: Learning and response-conflict resolution in the dimensional change card sort task. *Child Development, 84*(4), 1308–1323. https://doi.org/10.1111/cdev.12044.
Rennie, D. A., Bull, R., & Diamond, A. (2004). Executive functioning in preschoolers: Reducing the inhibitory demands of the dimensional change card sort task. *Developmental Neuropsychology, 26*(1), 423–443. https://doi.org/10.1207/s15326942dn2601_4.
Towse, J. N., Redbond, J., Houston-Price, C. M. T., & Cook, S. (2000). Understanding the dimensional change card sort: Perspectives from task success and failure. *Cognitive Development, 15*(3), 347–365. https://doi.org/10.1016/S0885-2014(00)00021-6.
Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46.
Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science, 7*(4), 121–126.
Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.). *Blackwell handbook of childhood cognitive development* (pp. 445–469). Malden, MA, US: Blackwell Publishers Ltd.
Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. A. Nelson, & M. Luciana (Eds.). *Handbook of developmental cognitive neuroscience* (pp. 553–574). Cambridge, MA, US: MIT Press.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1.
Zelazo, P. D., Reznick, J. S., & Piñon, D. E. (1995). Response control and the execution of verbal rules. *Developmental Psychology, 31*(3), 508–517. https://doi.org/10.1037/0012-1649.31.3.508.