Title

Training Children's Understanding of Numbers Using Compositional Cues

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Vilches, Lucia A
Srinivasan, Akshita
Liu, Jiaxin
et al.

Publication Date

2026

Training Children’s Understanding of Numbers Using Compositional Cues

Lucia Vilches (lvilches@g.harvard.edu)[1], Jiaxin Liu[1], Akshita Srinivasan[2], & Elizabeth Spelke[1]
[1]Department of Psychology, Harvard University
[2]Department of Psychology, Stanford University

Abstract

Infants form exact representations of small quantities (1 to 3) but only approximate representations of larger quantities. How do children acquire exact large number concepts? We tested whether training showing how sets of 5 or 6 can be decomposed into smaller sets increases children’s understanding of *five* and *six*. We compared compositional training, using expressions such as “I have five pets: three dogs and two cats” paired with visual displays of the subsets, to training that drew attention to each object in succession, as in counting (“I have five pets, here’s one, and another, and another, and another, and another.”). Children who did not know the trained number word at the outset performed significantly better at test in the compositional condition. Thus, children can leverage their knowledge of small numbers to build representations of larger numbers.

Keywords: Compositionality; Number Concepts

Introduction

From infancy, humans and non-human animals are equipped with nonsymbolic numerical abilities: the parallel individuation (PI) system, which supports exact representations of 1–3 objects, and the approximate number system (ANS), which supports approximate representations of numbers beyond the limits of PI (e.g., the distinction between 20 and 40; see Carey, 2009; Dehaene, 1997; Feigenson et al., 2004; Spelke, 2011; Spelke, 2022, for reviews). Because neither system supports exact representations of numbers beyond 3, here we ask how children come to represent larger numbers.

A variety of proposals have aimed to address this question. First, children may be innately endowed with the concepts “1”, “+”, and “=”, as well as a recursive rule—the successor function—which combines them to generate the full range of natural number concepts. On this view, children need only learn the number words and their order (Leslie et al., 2008).

Second, number concepts may be bootstrapped from a culture-specific counting routine (Carey, 2009), as children struggle to learn the meanings of specific number words, even after they recite the count list and recognize that it is relevant to quantity (Wynn, 1990, 1992). Much of this research uses the “Give-N” task, in which a child is asked to give the experimenter *one* object, then, if they succeed, *two*, and so on, until the child either successfully gives all numbers in their count list or fails on two trials asking for a particular number. By this task, children first learn the meaning of *one*, then *two*, then *three,* but eventually make an inductive leap, inferring the correspondence between the ordering of the words in their count list and the ordering of the numerical magnitudes to which these words refer (Sarnecka & Carey, 2008; Wynn, 1990, 1992). Thus, children construct the logic of the natural number system as they master the logic of their counting procedure: After mapping the early number words to representations from the PI system and learning the ordering of the words in their count list, they discover an analogy between moving to the next number word in the count list and adding an object to a set, ultimately realizing that the last number word reached when counting a set represents its cardinal value (Carey, 2009).

This theory maps well onto the process by which children in cultures that emphasize counting learn the number words and the logic of counting, but cross-cultural research suggests that mastery of the count routine is not necessary for all forms of exact-number understanding. For example, Jara-Ettinger et al. (2016) found that Tsimane children come to understand exact numerical transformations without first mastering counting. Another challenge to Carey’s theory comes from Davidson et al. (2012), who found that even cardinal-principle knowers show limited understanding of the logic of counting, calling into question whether children make a semantic induction over the number words in their counting range.

Other investigators propose that children acquire exact large-number concepts by composing their representations of small numbers (Spelke, 2017; Dehaene et al., 2025). On Spelke’s account, this process depends on core knowledge systems and the combinatorial resources of natural languages, and it occurs in four steps. In the first three steps, children learn the first three number words and the natural language features that later allow them to learn larger numbers: noun phrases (e.g., *a cup, the cat, your hand*) that link to representations of objects, agents, and visual forms. Learning these noun phrases allows them to construct a system of representations of object kinds.

Next, children learn expressions that refer to two or three sets of individuals of different kinds or with different identities (e.g., *the dog and the cat; my cup and your cup*), including phrases with nouns at different taxonomic levels (e.g., *Look at those birds: there’s a duck, a goose, and a swan*). Then, children use their knowledge of noun phrases, object kinds, and the first three number words to construct representations of larger numbers. For example, by mastering expressions such as, “John has two pets: a dog and a cat”, one can interpret expressions such as “John has five pets: three dogs and two cats.” Hearing such phrases, children may learn

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

words for larger numbers by addition or multiplication, constructing representations of exact large numbers from representations of exact smaller numbers.

A recent proposal by Dehaene et al. (2025) also focuses on the compositional structure of number but locates the source of this combinatorial capacity in an unlearned, internal language of thought (LoT) rather than in mastery of a natural language. Dehaene and colleagues (2025) note that Spelke's proposal leaves the emergence of such linguistic resources unaddressed. According to Dehaene and colleagues, exact number concepts originate in an internal LoT that supports the recursive composition of sets. The same representational primitives that allow humans to construct and recognize structured geometric patterns (repetition, concatenation, and recursive embedding) are applied to sets of objects, yielding intuitive representations of how integers combine by addition and multiplication. These operations generate exact representations of natural numbers from a minimal set of primitives, without presupposing prior access to a linguistic numeral system. A central component of this account is the principle of minimum description length (MDL): the cognitive complexity of a number is proportional to the length of its shortest internal description. Empirical support for this proposal comes from an analysis of lexical databases from six languages, relating number word frequency to the compositional complexity of the corresponding number concepts (Pajot, Sablé-Meyer, & Dehaene, 2025).

Dehaene et al.'s tests of their proposal focus primarily on patterns in adult language use, but research provides evidence that children also are sensitive to the compositional structure of sets. When larger sets appear in groups of smaller, subitizable subsets, children use the grouping cues to construct exact representations of larger numbers (Starkey & McCandliss, 2014; Feigenson & Halberda, 2004). Here we ask whether compositional training supports young children's learning of the meanings of words for such quantities.

Methods

We trained children on *five* or *six* either with compositional language, such as "five pets: three cats and two dogs," paired with visual images of the relevant subsets, or with a control procedure that, like counting, labeled a set of 5 or 6 objects by calling attention to the individual objects in succession. If children can use small-number meanings to learn larger number words, then children given compositional training should perform better on tests of the numbers they trained on than their counterparts given control training. Our methods and analysis plan were preregistered before data collection on OSF: https://osf.io/ux5ea/overview?view_only=87d564781ef04da28b82c4ca3d704af1

Participants

Ninety-two typically developing children contributed data to this study (mean age = 46 months; range = 30.73–53.73; 46 girls, 46 boys). All participants were monolingual English speakers (≥90% English exposure by caregiver report), recruited through a lab database. Our sample size was determined via power analysis in G*Power, assuming a medium effect size (d = 0.6) and 80% power, and was increased slightly to ensure even assignment across conditions. An additional 26 children began the study but were excluded per preregistered criteria.

Design and Procedure

Data collection occurred over video conferencing. Caregivers were present to set up the video screen and accompanied the child throughout the experiment. They were instructed not to help their child or remind them of strategies they could use.

All children completed three phases in a fixed order: a pretest (Point-to-N) assessing baseline knowledge of the number words *one* through *six*, a training phase (compositional or control), and a test phase assessing their understanding of the meanings of the words *five* and *six.*

Participants were assigned to one of two training conditions: compositional training (n = 46) or control training (n = 46) of either *five* (n = 23) or *six* (n = 23) in a 2 × 2 design crossing training condition with trained number.

The Point-to-N pretest, compositional and control training, and assessment are described in detail below.

Point-to-N. To equate children's number word knowledge across the training conditions, all children completed this task at the beginning of their session. We administered the Point-to-N task via video conference, using a titration procedure instead of Wynn's (1992) "Give-N" task, because piloting showed that children looked to their parents for help when the Give-N task was administered at home.

On each trial, children saw four plates of different colors, each displaying a different number of identical objects. Different kinds of objects appeared on different trials. The experimenter asked the child to identify the plate containing a given number of items (e.g., "Which plate has two candies?"). Children responded either by naming the color of a plate or by pointing. In the latter case, caregivers reported the color of the plate that the child pointed to.

The test began with a trial requesting *one*. If a child selected the correct plate, the task continued up the count list in order with each correct response. If a child gave an incorrect response to a given number, the next trial asked again for the previous number. The task ended when a child either succeeded on every trial up to *six* or when they failed on a tested number twice.

We determined each child's knower level based on the highest number for which they consistently pointed to the correct plate. For example, a child who succeeded on trials for *one* and *two* but failed *three* twice was classified as a two-knower. While previous work has revealed no stable five- or six-knower stage (e.g., Carey, 2009; Lee & Sarnecka, 2010; Lee & Sarnecka, 2011; Sarnecka & Lee, 2009; Wynn, 1990, 1992), here we use "five-knower" and "six-knower" to refer to children who succeeded up to *five* or *six* on the Point-to-N task. Additionally, we use "non-knower" to refer to children who did not succeed on any trials on this task. We use these labels with no assumption that they refer to reliable stages.

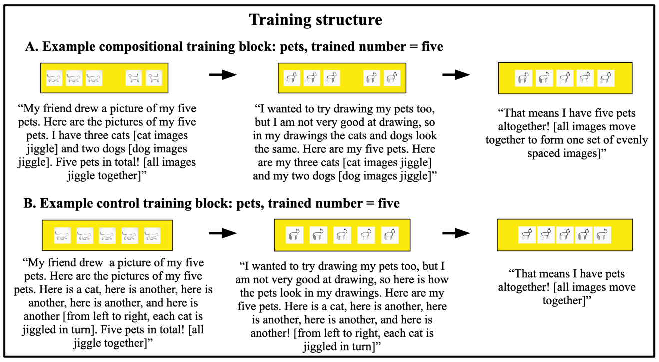

Figure 1: Example training blocks for each condition. Children completed three blocks: pets, berries, and drinks, each followed by two feedback-based practice trials. The figure shows the pets block for children trained on five.

Training. After completing the Point-to-N task, children were assigned to either the compositional or control training condition and to training on *five* or *six*. Knower level, age, and gender were considered during assignment to balance the number word knowledge and individual characteristics of the children in the two training conditions. Children received three training blocks presenting pets, berries, and drinks. Each block included an initial teaching phase followed by two practice trials with feedback, for a total of three teaching blocks and six practice trials. Below, each training condition is described using the training on *five*.

Compositional training. In this condition, children were presented with visual and linguistic cues to show how a set of five or six items can be composed from two smaller sets. In the first training block, the experimenter told the child, "My friend drew a picture of my five pets" while presenting an image of 3 cats and 2 dogs in a line with a space between the two kinds of pets. Then the child was told, "Here are the pictures of my five pets. I have three cats and two dogs, so I have five pets in total." The set of cats jiggled while the experimenter said, "three cats", and the set of dogs jiggled while the experimenter said, "two dogs." The pets jiggled all at once while the experimenter referred to the five pets together.

Because we wanted children to understand the set of pets as a single set of five, rather than as separate subsets of two and three, children were then shown new pictures of animals with identical shapes, sizes, and colors. These were animated as in the visually distinctive training display while the experimenter said, "These are my three cats" and "these are my two dogs," as each subset jiggled. The experimenter then stated that they had five pets total, as all five animals moved to form a single, equally spaced line.

Each block ended with two practice trials featuring three ungrouped sets varying in number from 4 to 7. The children were asked, "Which tray has five pets: three cats and two dogs?" After giving their answer, children received informative feedback. All children then saw the set of five pets separated spatially into sets of three and two, and the subsets of the correct set jiggled as the experimenter said, "See? There are three cats here and two dogs here." The same sequence was repeated in two additional blocks with different objects: five berries represented first as three raspberries and two blackberries, and five drinks represented first as three glasses of milk and two glasses of juice.

The compositional training on *six* was identical, but with three items in each subset.

Control training. The control condition presented displays featuring the same objects as those used in the compositional training condition, but all the objects looked the same, were spaced evenly, and were said to belong to a single set (e.g., five pets). The experimenter drew attention to each object individually, saying, "I have five/six pets. Here is a cat, here is another, here is another, here is another, and here is another." The objects jiggled one by one as the experimenter referred to each pet. The experimenter then stated that they had five pets total, as all five animals moved toward one another, forming a more compact single line.

As in the compositional condition, children in the control condition were given two practice trials after each of the three training blocks and received feedback after each one. In the first training block, the practice trials featured three different, spatially ungrouped sets of the ambiguous pet drawings with varying numbers, and children were asked which set had *five* pets. After the child selected a set, the experimenter provided corrective feedback. The images in the correct set moved to spread apart. Then, as the experimenter referred to each object, the object was jiggled individually. As in the compositional condition, the training was repeated twice with different objects (berries and drinks).

The training on *six* was identical, but with six items.

Assessment. After this training, all children completed 16 test trials to assess their understanding of the words *five* and *six*. Children viewed three linear arrays of objects varying in number from 4 to 7, one of which presented either 5 or 6 objects. Children were tested on both number words, regardless of their training condition. Four trial types were given, featuring: visual grouping cues (the objects in the set were spatially grouped into subsets of 2 and 3 or 3 and 3), compositional language (e.g., "one of these sets has five cookies: three are chocolate chip and two are oatmeal raisin"), both cues, or neither cue. Trials without visual grouping displayed evenly spaced items, and trials without compositional language used only the cardinal label (e.g., "Which tray has five cookies?"). Each of the four trial types was tested 4 times for a total of 16 trials.

On each trial, children were shown three sets of a particular object category (e.g., cookies, flowers, hats), each containing a different number of items. The object categories used during the assessment were distinct from those used during training. Children were asked to identify the set that matched a verbal prompt referencing the number *five* or *six* (e.g., "Which tray has exactly five cookies?"). Children responded by saying the color of the tray they thought was the correct answer or by pointing to their answer, conveyed by their parent. The position of the correct set varied across trials. The trials were presented in a pseudo-random order such that no more than three trials of the same type or same target number appeared consecutively.

Analysis

Accuracy on each assessment trial was coded as correct (1) or incorrect (0). All analyses use mixed-effects logistic regression models with a random intercept for participants to account for repeated trials within children. Age in months was mean-centered and included as a continuous covariate where specified. Training condition was coded as control versus compositional. Test trial type was treated as a categorical fixed effect with trials containing both compositional language and visual grouping as the reference level. Target type was coded as trained number versus untrained number. When knower level was used as a covariate, it was entered as a mean-centered numeric predictor; when used to define groups, it was treated categorically.

We first conducted the preregistered analyses. The primary preregistered model predicted trial-level accuracy from training condition and age, with a random intercept for participant. A second preregistered model added test trial type as an additive categorical predictor to test whether accuracy differed across the four assessment trial types while estimating the overall effect of training condition.

Because the preregistered analyses included children who had demonstrated knowledge of the number they were trained on before training began—an inclusion that could artificially inflate our measures of children's learning—we conducted follow-up analyses excluding data from such children. Children trained on *five* were included only if their pretest knower level was below five, and children trained on *six* were included only if their pretest knower level was below six. All further analyses were conducted using this subset of our sample.

Within this subset, we reran the primary preregistered model, predicting accuracy across all assessment trials from training condition and age. We then repeated this model with pretest knower level included as a covariate to assess whether training condition predicted accuracy after accounting for baseline number word knowledge. We also fit the same condition model to trials testing the trained number only.

We next asked whether the effect of training depended on children's knowledge of the two small number words used to teach the target number word in the compositional training condition. Children who were three-knowers or above were coded as knowing both component number words, and children below the three-knower level were coded as not knowing those words. We fit a model including training condition, small number knowledge group, their interaction, and the relevant covariates.

Finally, we asked whether the effects of compositional training were specific to the trained number or generalized to the untrained number. Because the training effect could depend on which compositional cues were available at test, this model included trial type and target type (trained vs. untrained). The full model, therefore, included training condition, test trial type, target type, all interactions among these predictors, age, knower level, and a random intercept for participant. The primary test of interest was the interaction of Training Condition (compositional vs. control) × Trial Type (visual grouping and compositional language cues vs. only visual grouping cues vs. only compositional language cues vs. neither type of cue) and Target Type (trained vs. untrained number).

Results

The 92 children in the experiment spanned a wide range of knower levels, distributed across the experimental conditions. Because knower level was imperfectly balanced across training conditions and trained numbers (Table 1), we report follow-up models adjusting for knower level. Children ranged in age from 30.7 to 53.7 months ($M = 46.0$, $SD = 5.44$), with similar ages across conditions (control: $M = 46.4$, $SD = 5.03$; compositional: $M = 45.7$, $SD = 5.86$).

Table 1: Full sample composition.

Full sample composition by knower level, condition, and trained number

	Control			Compositional			
Knower level	5	6	Total	5	6	Total	Grand total
non-knower	1.00	2.00	**3.00**	1.00	1.00	**2.00**	**5.00**
1-knower	0.00	1.00	**1.00**	1.00	1.00	**2.00**	**3.00**
2-knower	2.00	6.00	**8.00**	3.00	4.00	**7.00**	**15.00**
3-knower	6.00	4.00	**10.00**	4.00	5.00	**9.00**	**19.00**
4-knower	3.00	3.00	**6.00**	3.00	4.00	**7.00**	**13.00**
5-knower	6.00	1.00	**7.00**	3.00	3.00	**6.00**	**13.00**
6-knower	5.00	6.00	**11.00**	8.00	5.00	**13.00**	**24.00**
N	**23.00**	**23.00**	**46.00**	**23.00**	**23.00**	**46.00**	**92.00**
Mean	**4.09**	**3.39**	**3.74**	**4.09**	**3.70**	**3.89**	**3.82**
SD	**1.59**	**1.97**	**1.81**	**1.86**	**1.74**	**1.79**	**1.79**
Variance	**2.54**	**3.89**	**3.26**	**3.45**	**3.04**	**3.21**	**3.21**

Our primary preregistered analyses asked whether compositional training improved overall test performance across the full sample. Although children in the compositional condition were numerically more accurate than children in the control condition, the model did not reveal a significant effect of condition, $OR = 1.56$, 95% CI [0.89, 2.75], $p = .123$. Age reliably predicted accuracy, $OR = 1.14$, 95% CI [1.08, 1.20], $p < .001$.

Our next preregistered analysis tested whether test performance varied by test trial type, while again testing the overall effect of training condition. Condition remained non-significant, $OR = 1.56$, 95% CI [0.89, 2.75], $p = .123$, and age remained a reliable predictor, $OR = 1.14$, 95% CI [1.08, 1.20], $p < .001$. No trial type differed significantly from the reference trial type, which included both compositional and visual grouping cues. Thus, in the full sample, we found no evidence that compositional training influenced children's use of the cues present at test.

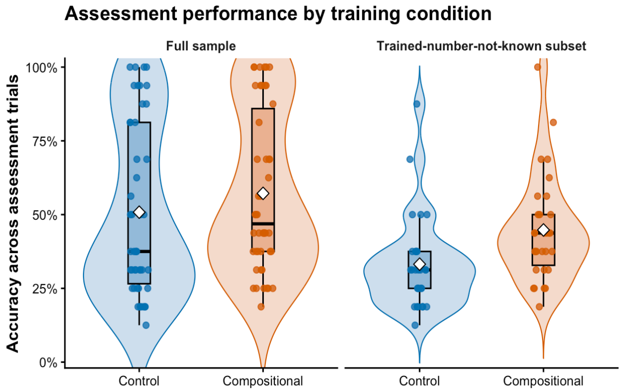

Figure 2: Assessment performance by training condition for the full sample and the trained-number-not-known subset.

The performance of the subset of children who did not already know the number word on which their training focused provides the most direct test of the primary question of interest. For these children, compositional training predicted higher accuracy across all assessment trials, relative to control training, OR = 1.74, 95% CI [1.21, 2.51], p = .003. The effect of training condition remained significant after adjusting for pretest knower level, OR = 1.63, 95% CI [1.15, 2.32], p = .006. Thus, training that composed a large number from small subsets led to greater gains in children's knowledge of the corresponding number word than did training that called attention to each member of the set in sequence. When this analysis was restricted to assessment trials asking for the trained number, rather than both numbers (5 and 6), children in the compositional condition remained numerically more accurate than children in the control condition, but the effect of condition was not significant, OR = 1.53, 95% CI [0.96, 2.45], p = .077.

Table 2: Composition of the subset of data, including only children who did not know the number they were trained on before training.

Trained-number-not-known subset by knower level, condition, and trained number

	Control			Compositional			
Knower level	**5**	**6**	**Total**	**5**	**6**	**Total**	**Grand total**
non-knower	1.00	2.00	**3.00**	1.00	1.00	**2.00**	**5.00**
1-knower	0.00	1.00	**1.00**	1.00	1.00	**2.00**	**3.00**
2-knower	2.00	6.00	**8.00**	3.00	4.00	**7.00**	**15.00**
3-knower	6.00	4.00	**10.00**	4.00	5.00	**9.00**	**19.00**
4-knower	3.00	3.00	**6.00**	3.00	4.00	**7.00**	**13.00**
5-knower	0.00	1.00	**1.00**	0.00	3.00	**3.00**	**4.00**
6-knower	0.00	0.00	**0.00**	0.00	0.00	**0.00**	**0.00**
N	**12.00**	**17.00**	**29.00**	**12.00**	**18.00**	**30.00**	**59.00**
Mean	**2.83**	**2.47**	**2.62**	**2.58**	**3.06**	**2.87**	**2.75**
SD	**1.11**	**1.37**	**1.27**	**1.24**	**1.39**	**1.33**	**1.29**
Variance	**1.24**	**1.89**	**1.60**	**1.54**	**1.94**	**1.77**	**1.68**

Next, we investigated whether the compositional training benefit depended on children's knowledge of the component number words used during training (i.e., *two* and *three*). Across all assessment trials, the Condition × Component Knowledge interaction was not significant, OR = 1.71, 95% CI [0.84, 3.51], p = .142; likelihood-ratio test, $\chi^2(1) = 2.13$, p = .144, but estimated marginal means showed the predicted pattern. Children who did not know the component number words showed no reliable benefit of compositional training relative to control, predicted accuracy = .298 versus .336, p = .551. In contrast, children who knew the component number words showed significantly higher accuracy in the compositional condition than in the control condition, predicted accuracy = .514 versus .342, p = .001. Thus, the benefit of compositional training was clearest among children who knew the meanings of *two* and *three*.

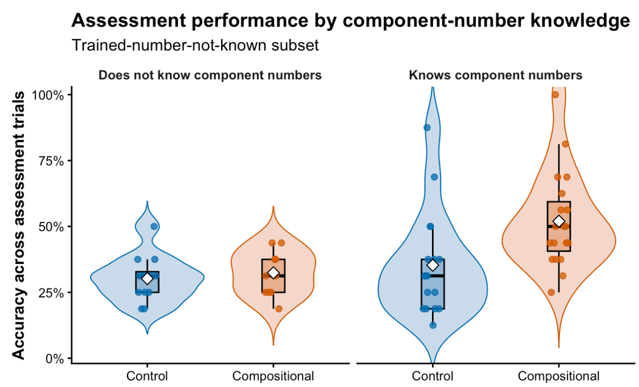

Figure 3: Assessment performance by component-number knowledge.

Our next model examined whether the condition effect varied based on which compositional cues were available at test: visual grouping and compositional language cues, only visual cues, only language cues, or neither cue. The primary test of interest was the Condition × Trial Type interaction, which showed no evidence that the condition effect differed across trial types, $\chi^2(3) = 1.35$, p = .716. As descriptive follow-ups, we estimated the condition effect separately within each trial type and found a compositional advantage on trials with both compositional language and visual grouping (p = .026), and on trials with only visual grouping cues (p = .049). A compositional advantage was not found for the other two trial types, suggesting that visual grouping cues are particularly beneficial for children in the compositional condition.

Because the overall interaction was not significant, these simple contrasts should not be interpreted as evidence that trial type reliably moderated the training effect. Indeed, when the same analysis was restricted to trials asking for the trained number, the Condition × Trial Type interaction remained non-significant, $\chi^2(3) = 3.55$, p = .314, and no trial-type-specific condition contrast was significant. Thus, although the descriptive contrasts were concentrated in trials with visual grouping cues, the non-significant interaction provides no evidence that compositional training worked because it drew attention to cues present at test.

Because the condition effect in previous analyses was reliable when including all assessment trials, but weaker when analyses were restricted to trained-number trials, our final analyses directly tested whether the compositional benefit across trial types differed when children were tested on the number they were trained on (5 or 6) or on the other number. The Condition × Trial Type × Target Type interaction was not significant, $\chi^2(3) = 2.98$, $p = .395$. Thus, we found no evidence that the compositional-training effect was specific to the trained number, although untrained-number trials were harder than trained-number trials in the reference trial type, OR = 0.35, $p = .015$. It is possible that children used the knowledge they had gained about the number they were trained on to infer the meaning of the untrained number: given that *five* decomposes into *two* and *three,* they might have inferred that *six,* a number that exceeds *five* by one, will decompose into *three* and *three.* Further research is needed, however, to test this possibility.

In summary, our preregistered analyses, which included the full sample, found no significant effect of compositional training on assessment performance. However, when analyses were restricted to children who had not, at pretest, demonstrated knowledge of the number they were trained on, compositional training predicted higher assessment performance, even when adjusting for pretest knower level. This benefit was clearest among children who knew the component number words used in training, though the formal moderation was not significant. Analyses comparing performance on trained-number and untrained-number assessment trials did not provide clear evidence that the benefit was specific to the trained number, raising the possibility that compositional training has broader effects on children's learning.

Discussion

This study tested whether brief exposure to compositional expressions (e.g., "five pets: three cats and two dogs"), paired with grouped visual arrays, helps children learn the meaning of words for larger numbers. Across the full sample, compositional training produced no reliable improvement in children's number word learning when controlling for age. However, when we examined training effects only in children who did not already know the number they were trained on, compositional training predicted greater number word learning than control training.

These findings are consistent with compositional accounts of number-word learning. If children can use known small-number meanings to build representations of larger numbers, then training that explicitly presents a larger set as composed of smaller subsets should help children learn larger number words. The present data are consistent with this possibility, but they also suggest that compositional training was not equally effective for all children. The clearest benefit appeared among children who knew the meanings of the number words designating the subsets into which the trained number could be decomposed: *two* and *three*. Although the formal interaction between condition and knowledge of these small numbers was not significant, the estimated marginal means showed that children who knew the component number words showed a reliable compositional-training advantage, whereas children who did not know them did not.

Our findings bear on the role of language in the development of exact number concepts beyond 3 or 4. Because young children benefit from grouped, small subsets that are not named in studies of groupitizing and set composition (Starkey & McCandliss, 2014; Feigenson & Halberda, 2004), numerical language is not necessary for this development. Nevertheless, knowing the meanings of *two* and *three* may have allowed the children in the present experiment to treat *three cats* and *two dogs* as subsets that can be combined to support learning *five*. Our results favor this possibility, as compositional training was most effective for children who knew the small number words.

The trial-type analyses, however, qualify this conclusion: Children given compositional training showed higher predicted accuracy on both-cue trials and visual-only trials, but not on language-only or no-cue trials. On visual-only trials, the prompt included no words for the subsets, but visual grouping still enhanced children's performance in the compositional condition. Thus, it is likely that children learned the target number word through compositional training but relied on visual grouping cues to aid enumeration, as research on groupitizing has shown (Starkey & McCandliss, 2014).

The analyses comparing trained- to untrained-number trials raise further questions. If compositional training specifically taught children the meaning of the trained number word, the training effect should have been stronger for trained-number trials than for untrained-number trials. We did not find reliable evidence for this effect. Thus, the present data do not isolate the compositional-training benefit to the trained number itself. The benefit may reflect learning of the trained number, improved ability to exploit compositional structure at assessment, or some combination of these processes.

In sum, compositional training improved children's learning of words for exact numbers beyond four, especially among those who knew the smaller number words used in the compositional training: *two* and *three*. These findings provide evidence that children can use their knowledge of words for small exact numbers to learn words for larger exact numbers, as previously proposed (Dehaene et al., 2025; Spelke, 2017). The emergence of exact large-number concepts may therefore depend not only on exposure to number words or the count list, but also on the construction of large numbers from smaller ones. Because the analyses supporting these findings were not included in our preregistration, however, we view these conclusions as tentative, pending a full-scale replication of the present experiment.

References

Carey, S. (2009). The Origin of Concepts. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. https://doi.org/10.1016/j.cognition.2011.12.013

Dehaene, S. (1997). The number sense: How the mind creates mathematics. Oxford University Press.

Dehaene, S., Sablé-Meyer, M., & Ciccione, L. (2025). Origins of numbers: A shared language-of-thought for arithmetic and geometry? Trends in Cognitive Sciences, 29(6), 526–540. https://doi.org/10.1016/j.tics.2025.03.001

Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, *91*(2), 173–190. https://doi.org/10.1016/j.cognition.2003.09.003

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. Trends in Cognitive Sciences, 8(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2016). Mastery of the logic of natural numbers is not the result of mastery of counting: Evidence from late counters. Developmental Science, 20(6). https://doi.org/10.1111/desc.12459

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science*, *34*(1), 51–67. https://doi.org/10.1111/j.1551-6709.2009.01063.x

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modeling. *Cognition*, *120*(3), 391–402. https://doi.org/10.1016/j.cognition.2010.10.003

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, *12*(6), 213–218. https://doi.org/10.1016/j.tics.2008.03.004

Pajot, M., Sablé-Meyer, M., & Dehaene, S. (2025). The compositional nature of number concepts: Insights from number frequencies. Cognition, 263, 106213. https://doi.org/10.1016/j.cognition.2025.106213

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. Cognition, 108(3), 662–674. https://doi.org/10.1016/j.cognition.2008.05.007

Spelke, E. S. (2011). Chapter 18 - Natural Number and Natural Geometry. In *Space, Time and Number in the Brain* (pp. 287–317). Elsevier Inc. https://doi.org/10.1016/B978-0-12-385948-8.00018-9

Spelke, E. S. (2017). Core knowledge, language, and number. Language Learning and Development, 13(2), 147–170. https://doi.org/10.1080/15475441.2016.1263572

Spelke, E. (2022). What babies know: Core knowledge and composition: Volume 1. Oxford University Press.

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology*, *126*, 120–137. https://doi.org/10.1016/j.jecp.2014.03.006

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. Cognitive Psychology, 24(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P